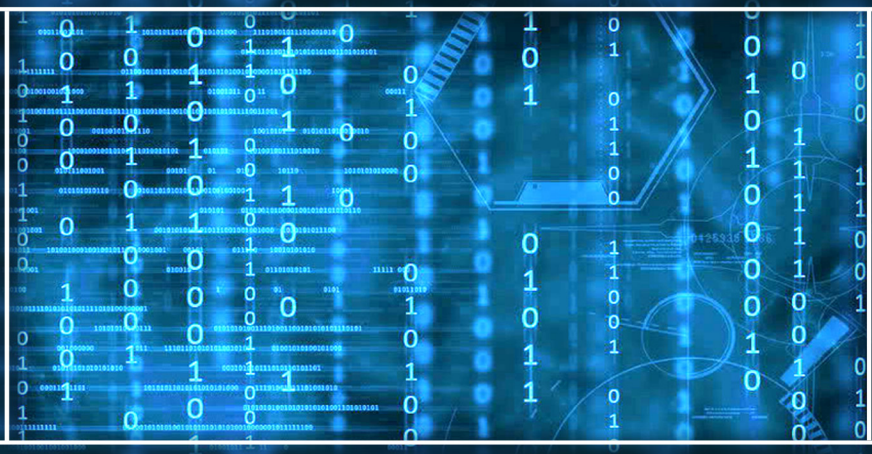
IJACSA
WHERE WISDOM SHARES
International Journal of Advanced Computer Science and Applications

Volume 15 Issue 10
October 2024



ISSN 2156-5570(Online)
ISSN 2158-107X(Print)

www.ijacsa.thesai.org

Editorial Preface

From the Desk of Managing Editor…

It may be difficult to imagine that almost half a century ago we used computers far less sophisticated than current home desktop computers to put a man on the moon. In that 50 year span, the field of computer science has exploded.

Computer science has opened new avenues for thought and experimentation. What began as a way to simplify the calculation process has given birth to technology once only imagined by the human mind. The ability to communicate and share ideas even though collaborators are half a world away and exploration of not just the stars above but the internal workings of the human genome are some of the ways that this field has moved at an exponential pace.

At the International Journal of Advanced Computer Science and Applications it is our mission to provide an outlet for quality research. We want to promote universal access and opportunities for the international scientific community to share and disseminate scientific and technical information.

We believe in spreading knowledge of computer science and its applications to all classes of audiences. That is why we deliver up-to-date, authoritative coverage and offer open access of all our articles. Our archives have served as a place to provoke philosophical, theoretical, and empirical ideas from some of the finest minds in the field.

We utilize the talents and experience of editor and reviewers working at Universities and Institutions from around the world. We would like to express our gratitude to all authors, whose research results have been published in our journal, as well as our referees for their in-depth evaluations. Our high standards are maintained through a double blind review process.

We hope that this edition of IJACSA inspires and entices you to submit your own contributions in upcoming issues. Thank you for sharing wisdom.

Thank you for Sharing Wisdom!

Kohei Arai
Editor-in-Chief
IJACSA
Volume 15 Issue 10 October 2024
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

CONTENTS

Paper 1: Development of an AI Based Failure Predictor Model to Reduce Filament Waste for a Sustainable 3D Printing Process
Authors: Noushin Mohammadian, Melissa Sofía Molina Silva, Giorgi Basiladze, Omid Fatahi Valilai
PAGE 1 – 6

Paper 2: Developing a Blockchain Based Supply Chain CO2 Footprint Tracking Framework Enabled by IoT
Authors: Mohammad Yaser Mofatteh, Roshanak Davallou, Chaida Ndahiro Ishimwe, Swaresh Suresh Divekar, Omid Fatahi Valilai
PAGE 7 – 12

Paper 3: Triggered Screen Restriction Framework: Transforming Gamified Physical Interventions
Authors: Majed Hariri, Richard Stone, Ulrike Genschel
PAGE 13 – 22

Paper 4: AI in the Detection and Prevention of Distributed Denial of Service (DDoS) Attacks
Authors: Sina Ahmadi
PAGE 23 – 29

Paper 5: A Hybrid Regression-Based Network Model for Continuous Face Recognition and Authentication
Authors: Bhanu Kiran Devisetty, Ayush Goyal, Avdesh Mishra, Mais W Nijim, David Hicks, George Toscano
PAGE 30 – 42

Paper 6: Designing Conversational Agents for Student Wellbeing
Authors: Jieyu Wang, Li Zhang, Dingfang Kang, Katherina G. Pattit
PAGE 43 – 52

Paper 7: Data Encoding with Generative AI: Towards Improved Machine Learning Performance
Authors: Abdelkrim SAOUABE, Hicham OUALLA, Imad MOURTAJI
PAGE 53 – 57

Paper 8: Cross-Modal Hash Retrieval Model for Semantic Segmentation Network for Digital Libraries
Authors: Siyu Tang, Jun Yin
PAGE 58 – 66

Paper 9: AI-Powered Waste Classification Using Convolutional Neural Networks (CNNs)
Authors: Chan Jia Yi, Chong Fong Kim
PAGE 67 – 75

Paper 10: Attention Mechanism-Based CNN-LSTM for Abusive Comments Detection and Classification in Social Media Text
Authors: BalaAnand Muthu First, Kogilavani Shanmugavadive, Veerappampalayam Easwaramoorthy Sathishkumar, Muthukumaran Maruthappa, Malliga Subramanian, Rajermani Thinakaran
PAGE 76 – 85

Paper 11: Knowledge Graph-Based Badminton Tactics Mining and Reasoning for Badminton Player Training Pattern Analysis and Optimization
Authors: Xingli Hu, Jiangtao Li, Ren Cai
PAGE 86 – 94

Paper 12: Analysis and Usability Evaluation of Virtual Reality in Cultural Landscape Promotion Platform Application
Authors: Yufang Huang, Yin Luo, Jingyu Zheng, Xunxiang Li
PAGE 95 – 103

Paper 13: Migrating from Monolithic to Microservice Architectures: A Systematic Literature Review
Authors: Hossam Hassan, Manal A. Abdel-Fattah, Wael Mohamed
PAGE 104 – 116

Paper 14: A Review on NS Beyond 5G: Techniques, Applications, Challenges and Future Research Directions
Authors: Cui Zhiyi, Azana Hafizah Mohd Aman, Faizan Qamar
PAGE 117 – 128

Paper 15: Real Time Object Detection for Sustainable Air Conditioner Energy Management System
Authors: Chang Shi Ying, C. PuiLin
PAGE 129 – 138

Paper 16: The Review of Malaysia Digital Health Service Mobile Applications' Usability Design
Authors: Kah Hao Lim, Chia Yean Lim, Anusha Achuthan, Chin Ernst Wong, Vina Phei Sean Tan
PAGE 139 – 146

Paper 17: Badminton Tracking and Motion Evaluation Model Based on Faster RCNN and Improved VGG19
Authors: Jun Ou, Chao Fu, Yanyun Cao
PAGE 147 – 158

Paper 18: Cyberbullying Detection on Social Networks Using a Hybrid Deep Learning Architecture Based on Convolutional and Recurrent Models
Authors: Aigerim Altayeva, Rustam Abdrakhmanov, Aigerim Toktarova, Abdimukhan Tolep
PAGE 159 – 167

Paper 19: Optimization with Adaptive Learning: A Better Approach for Reducing SSE to Fit Accurate Linear Regression Model for Prediction
Authors: Vijay Kumar Verma, Umesh Banodha, Kamlesh Malpani
PAGE 168 – 173

Paper 20: Efficient Remote Health Monitoring Using Deep Learning and Parallel Systems
Authors: Zakaria El Khadiri, Rachid Latif, Amine Saddik
PAGE 174 – 186

Paper 21: Smart Muni Platform: Efficient Emergency and Citizen Security Management Based on Geolocation, Technological Integration and Real Time Communication
Authors: Gerson Castro-Chucan, Hiroshi Chalco-Peñafiel, Norka Bedregal-Alpaca, Victor Cornejo-Aparicio
PAGE 187 – 195

Paper 22: Serious Games Model for Higher-Order Thinking Skills in Science Education
Authors: Siti Norliza Awang Noh, Hazura Mohamed, Nor Azan Mat Zin
PAGE 196 – 201

Paper 23: Lung CT Image Classification Algorithm Based on Improved Inception Network
Authors: Qianlan Liu
PAGE 202 – 213

Paper 24: Multilevel Characteristic Weighted Fusion Algorithm in Domestic Waste Information Classification
Authors: Min Li
PAGE 214 – 223

Paper 25: Novel Biomarkers for Colorectal Cancer Prediction
Authors: Mohamed Ashraf, M. M. El-Gayar, Eman Eldaydamony
PAGE 224 – 236

Paper 26: Environmental and Economic Benefit Analysis of Urban Construction Projects Based on Data Envelopment Analysis and Simulated Annealing Algorithm
Authors: Jie Gong
PAGE 237 – 246

Paper 27: Improving the Accuracy of Chili Leaf Disease Classification with ResNet and Fine-Tuning Strategy
Authors: Sayuti Rahman, Rahmat Arief Setyadi, Asmah Indrawati, Arnes Sembiring, Muhammad Zen
PAGE 247 – 255

Paper 28: Exploring the Impact of User Experience Elements on Virtual Reality for Emotion Regulation Through mVR-Real App
Authors: Irna Hamzah, Ely Salwana, Nilufar Baghaei
PAGE 256 – 263

Paper 29: Enhancing Emotion Regulation Through Virtual Reality Design Framework for Social-Emotional Learning (VRSEL)
Authors: Irna Hamzah, Ely Salwana, Nilufar Baghaei, Mark Billinghurst, Azhar Arsad
PAGE 264 – 272

Paper 30: Bridging the Gap: Machine Learning and Vision Neural Networks in Autonomous Vehicles for the Aging Population
Authors: Shengsheng Tan
PAGE 273 – 281

Paper 31: Machine Learning for Predicting Intradialytic Hypotension: A Survey Review
Authors: Saeed Alqahtani, Suhuai Luo, Mashhour Alanazi, Kamran Shaukat, Mohammed G Alsubaie, Mohammad Amer
PAGE 282 – 293

Paper 32: An Interactive Attention-Based Approach to Document-Level Relationship Extraction
Authors: Zhang Mei, Zhao Zhongyuan, Xu Zhitong
PAGE 294 – 301

Paper 33: Computer Modeling of the Stress-Strain State of Two Kvershlags with a Double Periodic System of Slits Weighty Elastic Transtropic Massif
Authors: Tursinbay Turymbetov, Gulmira Tugelbaeva, Baqlan Kojahmet, Bekzat Kuatbekov, Serzhan Maulenov, Bakhytzhan Turymbetov, Mukhamejan Abdibek
PAGE 302 – 308

Paper 34: Smart System for Driver Behavior Prediction
Authors: Hajar LAZAR, Zahi JARIR
PAGE 309 – 316

Paper 35: Development of Intelligent Learning Model Based on Ant Colony Optimization Algorithm
Authors: Xiaojing Guo, Xiaoying Zhu, Lei Liu

PAGE 317 – 327

Paper 36: Automated Detection of Malevolent Domains in Cyberspace Using Natural Language Processing and Machine Learning

Authors: Saleem Raja Abdul Samad, Pradeepa Ganesan, Amna Salim Al-Kaabi, Justin Rajasekaran, Singaravelan M, Peerbasha Shebbeer Basha

PAGE 328 – 341

Paper 37: Enhanced IVIFN-ExpTODIM-MABAC Technique for Multi-Attribute Group Decision-Making Under Interval-Valued Intuitionistic Fuzzy Sets

Authors: Bin Xie

PAGE 342 – 354

Paper 38: CoCoSo Framework for Management Performance Evaluation of Teaching Services in Sports Colleges and Universities with Euclidean Distance and Logarithmic Distance

Authors: Feng Li, Yuefei Wen

PAGE 355 – 364

Paper 39: Enhanced Methodology for Production-Education Integration and Quality Evaluation of Rural Vocational Education Under Rural Revitalization with 2-Tuple Linguistic Neutrosophic Numbers

Authors: Xingli Wang

PAGE 365 – 377

Paper 40: A Deep Learning Based Detection Method for Insulator Defects in High Voltage Transmission Lines

Authors: Wang Tingyu, Sun Xia, Liu Jiaxing, Zhang Yue

PAGE 378 – 384

Paper 41: A Review of Personalized Recommender System for Mental Health Interventions

Authors: Idayati Binti Mazlan, Noraswaliza Abdullah, Norashikin Ahmad, Siti Zaleha Harun

PAGE 385 – 393

Paper 42: Intelligent Service Book Sorting in University Libraries Based on Linear Discriminant Analysis Method

Authors: Changjun Wang, Fengxia You, Yu Wang

PAGE 394 – 403

Paper 43: Prediction of Booking Trends and Customer Demand in the Tourism and Hospitality Sector Using AI-Based Models

Authors: Siham Rekiek, Hakim Jebari, Kamal Reklaoui

PAGE 404 – 412

Paper 44: Optimizing Production in Reconfigurable Manufacturing Systems with Artificial Intelligence and Petri Nets

Authors: Salah Hammedi, Jalloul Elmelliani, Lotfi Nabli, Abdallah Namoun, Meshari Huwaytim Alanazi, Nasser Aljohani, Mohamed Shili, Sami Alshmrany

PAGE 413 – 425

Paper 45: TSO Algorithm and DBN-Based Comprehensive Evaluation System for University Physical Education

Authors: Yonghua Yang

PAGE 426 – 435

Paper 46: Robust Image Tampering Detection and Ownership Authentication Using Zero-Watermarking and Siamese Neural Networks

Authors: Rodrigo Eduardo Arevalo-Ancona, Manuel Cedillo-Hernandez, Francisco Javier Garcia-Ugalde

PAGE 436 – 446

Paper 47: Eating Behavior and Level of Knowledge About Healthy Eating Among Gym Users: A Multinomial Logistic Regression Study
Authors: Ana Huamani-Huaracca, Sebastián Ramos-Cosi, Michael Cieza-Terrones, Gina León-Untiveros, Alicia Alva Mantari
PAGE 447 – 455

Paper 48: Development of a Causal Model of Post-Millennials' Willingness to Disclose Information to Online Fashion Businesses (Thailand)
Authors: Apiwat Krommuang, Jinnawat Kasisuwan
PAGE 456 – 463

Paper 49: Facial Expression Classification System Using Stacked CNN
Authors: Aditya Wikan Mahastama, Edwin Mahendra, Antonius Rachmat Chrismanto, Maria Nila Anggia Rini, Andhika Galuh Prabawati
PAGE 464 – 472

Paper 50: Optimization of 3D Coverage Layout for Multi-UAV Collaborative Lighting in Emergency Rescue Operations
Authors: Dan Jiang, Rui Yan
PAGE 473 – 484

Paper 51: Analyzing the Impact of Occupancy Patterns on Indoor Air Quality in University Classrooms Using a Real-Time Monitoring System
Authors: Sri Ratna Sulitiyanti, Muhamad Komarudin, F. X Arinto Setyawan, Hery Dian Septama, Titin Yulianti, M. Farid Ammar
PAGE 485 – 492

Paper 52: Active Semi-Supervised Clustering Algorithm for Multi-Density Datasets
Authors: Walid Atwa, Abdulwahab Ali Almazroi, Eman A. Aldhahr, Nourah Fahad Janbi
PAGE 493 – 500

Paper 53: MSMA: Merged Slime Mould Algorithm for Solving Engineering Design Problems
Authors: Khaled Mohammad Alhashash, Hussein Samma, Shahrel Azmin Suandi
PAGE 501 – 516

Paper 54: Research on Credit Card Fraud Prediction Model Based on GAN-DNN Imbalance Classification Algorithm
Authors: Qin Wang, Mary Jane C.Samonte
PAGE 517 – 523

Paper 55: Impact Analysis of Informatization Means Driven by Artificial Intelligence Technology on Visual Communication
Authors: Lei Ni
PAGE 524 – 533

Paper 56: Revolutionizing Rice Leaf Disease Detection: Next-Generation SMOREF-SVM Integrating Spider Monkey Optimization and Advanced Machine Learning Techniques
Authors: Avip Kurniawan, Tri Retnaningsih Soeprobowati, Budi Warsito
PAGE 534 – 543

Paper 57: Human Dorsal Hand Vein Segmentation Method Based on GR-UNet Model
Authors: Zhike Zhao, Wen Zeng, Kunkun Wu, Xiaocan Cui
PAGE 544 – 552

Paper 58: A Proposed Batik Automatic Classification System Based on Ensemble Deep Learning and GLCM Feature Extraction Method
Authors: Luluk Elvitaria, Ezak Fadzrin Ahmad Shaubari, Noor Azah Samsudin, Shamsul Kamal Ahmad Khalid, Salamun, Zul Indra
PAGE 553 – 561

Paper 59: A Comprehensive Crucial Review of Re-Purposing DNN-Based Systems: Significance, Challenges, and Future Directions
Authors: Yaser M Al-Hamzi, Shamsul Bin Sahibuddin
PAGE 562 – 582

Paper 60: A Machine Learning Operations (MLOps) Monitoring Model Using BI-LSTM and SARSA Algorithms
Authors: Zeinab Shoieb Elgamal, Laila Elfangary, Hanan Fahmy
PAGE 583 – 593

Paper 61: Deep Learning Approach in Complex Sentiment Analysis: A Case Study on Social Problems
Authors: Bambang Nurdewanto, Kukuh Yudhistiro, Dani Yuniawan, Himawan Pramaditya, Mochammad Daffa Putra Karyudi, Yulia Natasya Farah Diba Arifin, Puput Dani Prasetyo Adi
PAGE 594 – 604

Paper 62: Evaluating the Effectiveness of the Binary PSO Method in Feature Selection to Improve the Detection of Android Botnets
Authors: Peng WANG, Zhijun WANG
PAGE 605 – 612

Paper 63: The Impact of the GQM Framework on Software Engineering Exam Outcomes
Authors: Reem Abdulaziz Alnanih
PAGE 613 – 622

Paper 64: Critical Success Factors of Microservices Architecture Implementation in the Information System Project
Authors: Mochamad Gani Amri, Teguh Raharjo, Anita Nur Fitriani, Nurman Rasyid Panusunan Hutasuhut
PAGE 623 – 633

Paper 65: Application Citespace Visualization Tool to Online Public Opinion Group Label: Generation, Dissemination and Trends
Authors: Jingyi Ju
PAGE 634 – 646

Paper 66: Individual Cow Identification Using Non-Fixed Point-of-View Images and Deep Learning
Authors: Yordan Kalmukov, Boris Evstatiev, Seher Kadirova
PAGE 647 – 657

Paper 67: A Modified Lightweight DeepSORT Variant for Vehicle Tracking
Authors: Ayoub El-alami, Younes Nadir, Khalifa Mansouri
PAGE 658 – 664

Paper 68: Multi-Site Cross Calibration on the LAPAN-A3/IPB Satellite Multispectral Camera with One-Dimensional Kalman Filter Optimization
Authors: Sartika Salaswati, Adhi Harmoko Saputro, Wahyudi Hasbi, Deddy El Amin, Patria Rachman Hakim, Silmie Vidiya Fani, Agung Wahyudiono, Ega Asti Anggari
PAGE 665 – 675

Paper 69: Combining BERT and CNN for Sentiment Analysis A Case Study on COVID-19
Authors: Gunjan Kumar, Renuka Agrawal, Kanhaiya Sharma, Pravin Ramesh Gundalwar, Aqsa kazi, Pratyush Agrawal, Manjusha Tomar, Shailaja Salagrama
PAGE 676 – 686

Paper 70: Visual Translation of Auspicious Beliefs in Quanzhou Xi Culture from the Perspective of Man-Machine Collaboration
Authors: Li Zheng, Xu Zhang, Huiling Guo
PAGE 687 – 698

Paper 71: Strength Calculation Method of Agricultural Machinery Structure Using Finite Element Analysis
Authors: Jing Yang
PAGE 699 – 706

Paper 72: Constructing Knowledge Graph in Blockchain Teaching Program Using Formal Concept Analysis
Authors: Madina Mansurova, Assel Ospan, Dinara Zhaisanova
PAGE 707 – 720

Paper 73: Analysis of Influencing Factors of Tourist Attractions Accessibility Based on Machine Learning Algorithm
Authors: Na Liu, Hai Zhang
PAGE 721 – 727

Paper 74: A Smart Contract Approach for Efficient Transportation Management
Authors: Abdullah Alshahrani, Ayman Khedr, Mohamed Belal, Mohamed Saleh
PAGE 728 – 738

Paper 75: Enhancing Educational Outcomes Through AI Powered Learning Strategy Recommendation System
Authors: Daminda Herath, Chanuka Dinuwan, Charith Ihalagedara, Thanuja Ambegoda
PAGE 739 – 748

Paper 76: Implementation of Lattice Theory into the TLS to Ensure Secure Traffic Transmission in IP Networks Based on IP PBX Asterisk
Authors: Olga Abramkina, Mubarak Yakubova, Tansaule Serikov, Yenlik Begimbayeva, Bakhodyr Yakubov
PAGE 749 – 756

Paper 77: Energy Optimization Management Scheme for Manufacturing Systems Based on BMAPPO: A Deep Reinforcement Learning Approach
Authors: Zhe Shao
PAGE 757 – 765

Paper 78: Design Science Research: Applying Integrated Fogg Persuasive Frameworks to Validate Rural ICT Design Requirements
Authors: Noela Jemutai Kipyegen, Benard Okelo
PAGE 766 – 777

Paper 79: Brain Tumor Segmentation of Magnetic Resonance Imaging (MRI) Images Using Deep Neural Network Driven Unmodified and Modified U-Net Architecture
Authors: Nunik Destria Arianti, Azah Kamilah Muda
PAGE 778 – 786

Paper 80: Indoor Landscape Design and Environmental Adaptability Analysis Based on Improved Fuzzy Control
Authors: Jinming Liu, Qian Hu, Pichai Sodbhiban
PAGE 787 – 795

Paper 81: Optimising Delivery Routes Under Real-World Constraints: A Comparative Study of Ant Colony, Particle Swarm and Genetic Algorithms
Authors: Rneem I. Aldoraibi, Fatimah Alanazi, Haya Alaskar, Abed Alanazi
PAGE 796 – 803

Paper 82: Volleyball Motion Analysis Model Based on GCN and Cross-View 3D Posture Tracking
Authors: Hongsi Han, Jinming Chang
PAGE 804 – 815

Paper 83: A DECOC-Based Classifier for Analyzing Emotional Expressions in Emoji Usage on Social Media
Authors: Shaya A. Alshaya
PAGE 816 – 822

Paper 84: A Smart IoT System for Enhancing Safety in School Bus Transportation
Authors: Yousef H. Alfaifi, Tareq Alhmiedat, Emad Alharbi, Ahad Awadh Al Grais, Maha Altalk, Abdelrahman Osman Elfaki
PAGE 823 – 831

Paper 85: Machine Learning Approaches Applied in Smart Agriculture for the Prediction of Agricultural Yields
Authors: Abourabia. Imade, Ounacer. Soumaya, Elghoumari. Mohammed yassine, Azzouazi. Mohamed
PAGE 832 – 843

Paper 86: K-Means and Morphology Based Feature Element Extraction Technique for Clothing Patterns and Lines
Authors: Xiaojia Ding
PAGE 844 – 850

Paper 87: Simulation Analysis of Obstacle Crossing Stability for Transmission Line Inspection Robot
Authors: Qianli Wang
PAGE 851 – 860

Paper 88: Precision Machining of Hard-to-Cut Materials: Current Status and Future Directions
Authors: Tengjiao CUI
PAGE 861 – 871

Paper 89: Analyzing VGG-19's Bias in Facial Beauty Prediction: Preference for Feminine Features
Authors: Nuno Fernandes, Sandra Soares, Joana Arantes
PAGE 872 – 881

Paper 90: Real-Time Self-Localization and Mapping for Autonomous Navigation of Mobile Robots in Unknown Environments
Authors: Serik Tolenov, Batyrkhan Omarov
PAGE 882 – 893

Paper 91: Optimizing LSTM-Based Model with Ant-Lion Algorithm for Improving Thyroid Prognosis
Authors: Maria Yousef
PAGE 894 – 902

Paper 92: Balancing Privacy and Performance: Exploring Encryption and Quantization in Content-Based Image Retrieval Systems
Authors: Mohamed Jafar Sadik, Noor Azah Samsudin, Ezak Fadzrin Bin Ahmad
PAGE 903 – 920

Paper 93: Tracking Computer Vision Algorithm Based on Fusion Twin Network
Authors: Xin Wang
PAGE 921 – 931

Paper 94: Graph Neural Networks and Dominant Set Algorithms for Energy-Efficient Internet of Things Environments: A Review
Authors: Dezhi Liao, Xueming Huang
PAGE 932 – 941

Paper 95: Development of Traffic Light and Road Sign Detection and Recognition Using Deep Learning
Authors: Joseph M. De Guia, Madhavi Devaraj
PAGE 942 – 952

Paper 96: Smart X-Ray Geiger Data Logger: An Integrated System for Detection, Control, and Dose Evaluation
Authors: Lhoucine Ben Youssef, Abdelmajid Bybi, Hilal Drissi, El Ayachi Chater
PAGE 953 – 961

Paper 97: A Feature Map Adversarial Attack Against Vision Transformers
Authors: Majed Altoub, Rashid Mehmood, Fahad AlQurashi, Saad Alqahtany, Bassma Alsulami
PAGE 962 – 968

Paper 98: Backbone Feature Enhancement and Decoder Improvement in HRNet for Semantic Segmentation
Authors: HanLei Feng, TieGang Zhong
PAGE 969 – 979

Paper 99: Machine Learning Approach to Identify Promising Mountain Hiking Destinations Using GIS and Remote Sensing
Authors: Lahbib Naimi, Charaf Ouaddi, Lamya Benaddi, El Mahi Bouziane, Abdeslam Jakimi, Mohamed Manaouch
PAGE 980 – 988

Paper 100: Pioneering Granularity: Advancing Native Language Identification in Ultra-Short EAP Texts
Authors: Zhendong Du, Kenji Hashimoto
PAGE 989 – 995

Paper 101: Feature Creation to Enhance Explainability and Predictability of ML Models Using XAI
Authors: Waseem Ahmed
PAGE 996 – 1007

Paper 102: Secret Sharing as a Defense Mechanism for Ransomware in Cloud Storage Systems
Authors: Shuaib A Wadho, Sijjad Ali, Asma Ahmed A. Mohammed, Aun Yichiet, Ming Lee Gan, Chen Kang Lee
PAGE 1008 – 1018

Paper 103: Core Scheduler Task Duplication for Multicore Multiprocessor System
Authors: Aya A. Eladgham, Nesreen I. Ziedan, Ibrahim Ziedan
PAGE 1019 – 1026

Paper 104: Enhancing Skin Cancer Detection with Transfer Learning and Vision Transformers
Authors: Istiak Ahmad, Bassma Saleh Alsulami, Fahad Alqurashi
PAGE 1027 – 1034

Paper 105: Accurate Head Pose Estimation-Based SO(3) and Orientation Tokens for Driver Distraction Detection
Authors: Xiong Zhao, Sarina Sulaiman, Wong Yee Leng
PAGE 1035 – 1043

Paper 106: Predicting the Most Suitable Delivery Method for Pregnant Women by Using the KGC Ensemble Algorithm in Machine Learning
Authors: Pusarla Sindhu, Parasana Sankara Rao
PAGE 1044 – 1051

Paper 107: MH-LViT: Multi-path Hybrid Lightweight ViT Models with Enhancement Training
Authors: Yating Li, Wenwu He, Shuli Xing, Hengliang Zhu
PAGE 1052 – 1061

Paper 108: Enhanced Fish Species Detection and Classification Using a Novel Deep Learning Approach
Authors: Musab Iqtait, Marwan Harb Alqaryouti, Ala Eddin Sadeq, Ahmad Aburomman, Mahmoud Baniata, Zaid Mustafa, Huah Yong Chan
PAGE 1062 – 1067

Paper 109: Reducing Traffic Congestion Using Real-Time Traffic Monitoring with YOLOv8
Authors: Sameerchand Pudaruth, Irfaan Mohammad Boodhun, Choo Wou Onn
PAGE 1068 – 1079

Paper 110: Enhancing Credit Card Fraud Detection Using a Stacking Model Approach and Hyperparameter Optimization
Authors: El Bazi Abdelghafour, Chrayah Mohamed, Aknin Noura, Bouzidi Abdelhamid
PAGE 1080 – 1087

Paper 111: Towards Interpretable Diabetic Retinopathy Detection: Combining Multi-CNN Models with Grad-CAM
Authors: Zakaria Said, Fatima-Ezzahraa Ben-Bouazza, Mounir Mekkour
PAGE 1088 – 1098

Paper 112: Breast Tumor Classification Using Dynamic Ultrasound Sequence Pooling and Deep Transformer Features
Authors: Mohamed A Hassanien, Vivek Kumar Singh, Mohamed Abdel-Nasser, Domenec Puig
PAGE 1099 – 1107

Paper 113: Classification of Moroccan Legal and Legislative Texts Using Machine Learning Models
Authors: Amina BOUHOUCHE, Mustapha ESGHIR, Mohammed ERRACHID
PAGE 1108 – 1114

Paper 114: ERCO-Net: Enhancing Image Dehazing for Optimized Detail Retention
Authors: Muhammad Ayub Sabir, Fatima Ashraf, Ahthasham Sajid, Nisreen Innab, Reem Alrowili, Yazeed Yasin
PAGE 1115 – 1122

Paper 115: Optimizing Text Summarization with Sentence Clustering and Natural Language Processing
Authors: Zahir Edress, Yasin Ortakci
PAGE 1123 – 1132

Paper 116: Hiding Encrypted Images in Audios Based on Cellular Automatas and Discrete Fourier Transform
Authors: Jose Alva Cornejo, Esdras D. Vasquez, Jose Calizaya Quispe, Roxana Flores-Quispe, Yuber Velazco-Paredes
PAGE 1133 – 1142

Paper 117: DBPF: An Efficient Dynamic Block Propagation Framework for Blockchain Networks
Authors: Osama Farouk, Mahmoud Bakrey, Mohamed Abdallah
PAGE 1143 – 1154

Paper 118: Skin Diseases Classification with Machine Learning and Deep Learning Techniques: A Systematic Review
Authors: Amina Aboulmira, Hamid Hrimech, Mohamed Lachgar
PAGE 1155 – 1173

Paper 119: An Investigation into the Risk Factors of Forest Fires and the Efficacy of Machine Learning Techniques for Early Detection
Authors: Asma Cherif, Sara Chaudhry, Sabina Akhtar
PAGE 1174 – 1184

Paper 120: Revolutionizing Historical Document Digitization: LSTM-Enhanced OCR for Arabic Handwritten Manuscripts
Authors: Safiullah Faizullah, Muhammad Sohaib Ayub, Turki Alghamdi, Toqeer Syed Ali, Muhammad Asad Khan, Emad Nabil
PAGE 1185 – 1194

Paper 121: A Secure Scheme to Counter the Man in the Middle Attacks in SDN Networks-Based Domain Name System
Authors: Frank Manuel Vuide Pangop, Miguel Landry Foko Sindjoung, Mthulisi Velempini
PAGE 1195 – 1201

Paper 122: Efficient Load-Balancing and Container Deployment for Enhancing Latency in an Edge Computing-Based IoT Network Using Kubernetes for Orchestration
Authors: Garrik Brel Jagho Mdemaya, Milliam Maxime Zekeng Ndadji, Miguel Landry Foko Sindjoung, Mthulisi Velempini
PAGE 1202 – 1210

Paper 123: Advanced Techniques for Optimizing Demand-Side Management in Microgrids Through Load-Based Strategies
Authors: Ramia Ouederni, Bechir Bouaziz, Faouzi Bacha
PAGE 1211 – 1219

Paper 124: Ensemble of Weighted Code Mixed Feature Engineering and Machine Learning-Based Multiclass Classification for Enhanced Opinion Mining on Unstructured Data
Authors: Ruchi Sharma, Pravin Shrinath
PAGE 1220 – 1230

Development of an AI Based Failure Predictor Model to Reduce Filament Waste for a Sustainable 3D Printing Process

Noushin Mohammadian[1], Melissa Sofía Molina Silva[2], Giorgi Basiladze[3], Omid Fatahi Valilai[4]*
School of Business, Social & Decision Science, Constructor University Bremen, Campus Ring 1, 28759, Bremen, Germany

***Abstract*—This paper delves into the integration of motion tracking technology for real-time monitoring in 3D printing, with a focus on the popular fused filament fabrication (FFF) technique. Despite FFF's cost-efficiency, prevalent printing errors pose significant challenges to its commercial and environmental viability. This study proposes a solution by incorporating motion tracking nodes into the 3D printing process, tracked by cameras, enabling dynamic identification and rectification of printing failures. Addressing key research questions, the paper explores the applicability of motion tracking for failure detection, its impact on printed object quality, and the potential reduction in 3D printing waste. The proposed real-time monitoring system aims to fill a critical gap in existing 3D printing procedures, providing dynamic failure identification. The study integrates machine learning, computer vision, and motion tracking technologies, employing an inductive theoretical development strategy with active learning iterations for continuous improvement. Highlighting the revolutionary potential of 3D printing and acknowledging challenges in continuous monitoring and waste management, the suggested system pioneers real-time monitoring, striving to enhance efficiency, sustainability, and adaptability to diverse production demands. As the study progresses into implementation, it aspires to contribute significantly to the evolution of 3D printing technologies.**

Keywords*—3D printing; Fused Filament Fabrication (FFF); motion tracking; environmental sustainability; printing waste reduction

I. Introduction

Since its creation in 1971, 3D printing has come a long way. However, it has just lately gained extreme popularity due to significant price reduction in both its equipment and materials. A technique of 3D printing known as fused filament fabrication (FFF) has grown in popularity making it the most widely utilized form of 3D printing to date [1]. This preeminence can be credited to its capacity to function without a heated print environment and possessing the ability to generate internal infills yielding a lightweight structure with enough support to have a strong shell [2]. Despite FFF cost-efficiency the prevalence of printing errors remains significant, presenting challenges to the economic and environmental viability of the approach. Consequently, many endeavors have been undertaken to mitigate printing failures. Previous studies have utilized cameras and image examination to detect commonly occurring errors such as “blocked nozzle” and “incomplete print” allowing the monitoring of both the external shape of the printed object and internal structure of its layers [1], [3].

Real-time monitoring remains a feature not widely incorporated into 3D printing processes meaning that technologies and techniques for this purpose are underdeveloped. Currently, the most effective means of rectifying issues during printing involves manual adjustment of printing parameters. This process requires extensive human experience and thus is not scalable to the industrial level [2].

Motion tracking has been applied in recent years to a wide range of fields such as videogame development and medical application [4], [5], [6]. This technology uses nodes tracked by cameras to determine movement in their axes. In this paper, we explore an approach to implement real-time monitoring by integrating motion tracking nodes into the printing process. These nodes will later be tracked by cameras to identify printing failures based on the movement or misplacement of the nodes within their coordinates. Our research questions guide this exploration: Can motion tracking be applied for detection of failures in 3D printing? Does incorporating 3D motion tracking points affect the quality of the printed object? Does the utilization of motion capture technology result in a noticeable reduction in 3D printing waste produced by failures?

By addressing these questions, we aim to shed light on the potential of motion tracking technology in revolutionizing real-time monitoring for enhanced 3D printing processes.

II. Literature Review

A. Adaptive 3D Printing Error Detection

The capacity to inspect objects in motion is central to various emerging additive manufacturing (AM) applications [7]. Through this work we are providing a novel approach for the automatic detection of a prints failure to reduce waste. The core component for the autonomous detection of failure in the system is a classification model which detects whether undesirable extrusion or deviations from the original CAD model exists [8].

By identifying errors as soon as they are produced the response rate of the system reaches or even surpasses the human reaction and the model can recognize inferior quality prints that humans will have a difficult time distinguishing with high accuracy. This self-diagnostic system has the potential to

*Corresponding author

be applied to other materials and manufacturing systems without human interaction [2].

B. AI Applications for Trajectory Interpretation and Correction

This research aims to employ physical camera systems and advanced image processing models to detect and identify errors occurring during the 3D printing process. These errors encompass phenomena such as printing misalignment, over-extrusion, and filament wastage. The primary objective of this endeavor is to establish a real-time monitoring mechanism that can promptly halt the 3D printing process in response to detected anomalies. To achieve this, we intend to integrate digital markers into the digital file intended for printing. Subsequently, the camera system will be tasked with the responsibility of capturing and analyzing the printing process to discern any deviations from the specified parameters, thereby facilitating immediate intervention when errors are detected.

Similar idea was described in research of Kucukdeger & Johnson: Prior to 3D scanning, three user-defined reference points (~1 mm diameter) were placed around the perimeter of the object using a black marker to provide reference points for the 3D printing frame. The object was then scanned using a calibrated single camera-projector structured light scanning system (HP 3D Structured Light Scanner Pro S2; HP) to acquire the point cloud data following our previously reported protocols [7].

C. Literature Gap Analysis

3D printers can quickly translate 3D design data into sophisticated forms by avoiding traditional cutting tools, fixtures, and many production phases, successfully tackling the difficulty of producing complex parts. As technology progresses, 3D printing has found widespread use in a variety of industries, including consumer electronics, automotive, aerospace, healthcare, defense, geographic mapping, and creative creation. For instance, BMW adopted 3D printing technology at some of its larger dealerships to produce spare parts for classic cars [9].

Since the first 3D printer was invented by Charles Hull in 1986. After that, a great achievement has been made in 3D printing [10]. In the year 2000, ZcorpÓ, an American company in collaboration with the RilecenÓ Institute, pioneered a color 3D printer utilizing inkjet printing technology. Toward the close of that year, an Israeli firm named Object Geometries introduced the Quadra 3D printer, which merged stereolithography apparatus (SLA) with 3D inkjet technology [11]. In 2010, a significant milestone was achieved in the field of medical 3D printing when an American firm known as OrganoxoÓ collaborated with an Australian company named InvetechÓ to produce human tissues and organs using live cells.

America in 2012, products in electronic area, automobile manufacturing, medical treatment and industrial machinery industry occupy a large share [10]. In this case, the manufacturer, BMWÓ, has invested in purchasing 3D printers and developing 3D designs. As a result, they were able to shut down its regular automobile production lines, which were only utilized to produce replacement parts. As a result, BMWÓ has adopted a make-to-order strategy, in which spare parts are produced only when there is a need for them.

The manufacturing industry is experiencing a significant transformation driven by the increasing demand for personalized products. 3D printing has emerged as a powerful tool, allowing for individualized manufacturing on an unprecedented scale [12]. This technology not only enhances product personalization but also revolutionizes industries by changing traditional production methods. The adoption of 3D printing results in faster production and reduced costs, enabling consumers to have more influence on the final product's design. Additionally, manufacturing facilities utilizing 3D printing are located closer to consumers, promoting a more flexible and responsive production process with enhanced quality control. Moreover, 3D printing reduces the need for global transportation, as manufacturing sites can be closer to end destinations, and efficient distribution can be achieved through advanced tracking technologies [13]. 3D printing presents several benefits, yet it also poses challenges, including the necessity for continuous product monitoring and the management of waste generated from filament usage. These challenges hinder its adoption across sectors, as implementing this technology entails risks without a guarantee of a positive return on investment in the production system. Despite these hurdles, 3D printing continues to evolve as a flexible and powerful technique in advanced manufacturing, marking its ongoing progress in the industry [13]. The result of literature study based on the defined themes is discussed in Table I.

TABLE I. LITERATURE STUDY BASED ON THE MAIN THEMES

Research Studies	Adaptive 3D printing error detection	AI models for analysis of deviations	inclusion of sustainability analysis for 3D printing waste
[14]	Using optical images for real-time defect detection through image correlation for temperature monitoring and thermal image analysis.	Both thermal images using FLIR Thermal Studio Pro for color gradient adjustment and optical image based on feature extraction	Mention of the advantage of in-process monitoring as a part of manufacturing because it provides the possibility of intervention or repair so that the print can be salvaged in this way preventing waste.
[15]	The pharmaceutical industry can utilize the 3D printing technology to print complex shapes to later determine their effectiveness in drug profile release by analyzing their volume and surface area.	By the utilization of machine learning algorithms, they were able to determine the prediction of drug release profile of FDM printed formulations	Machine Learning provides a suitable approach to modelling the 3D printing workflow. providing accuracies as high as 93% for values in the filament hot melt extrusion process in this way reducing the number of prints needed per trial.
[16]	The printer being utilized possess a Delta 3D printing Bed. The proposed approach has been evaluated on	Utilization of sensors to collect information of different variables in both the 3-axis velocity sensors and 3-axis	This work has proposed a new method for the construction of a fault detection model for 3D printers in this way they will be able to build a

	two cases in 3D printers: fault detection of 12 different join bearings, and fault detection in 3 synchronous bands.	angle sensors for error detection to test the proposed fCGAE technique to fault detection	deep feature space from raw signals to identify flaws and prevent them from happening in future prints.
[17]	This paper proposes a data-driven fault detection framework for semi-supervised scenarios where labeled training data from the system under consideration is imbalanced, but data from a related system (the "source") is readily available.	The goal of this paper is to develop an efficient fault detection algorithm for cyber-physical systems operating in scenarios where there is imbalanced labeled training data generated by the actual plant under consideration, but where training data from a related system is readily available.	This paper presents a domain-adaptation based technique capable of leveraging a small amount of labeled data generated by the real system together with data generated by an untuned simulator. The main result shows that optimization is possible to efficiently solve defects before they happen.
[18]	Solutions for the naturally occurring under-extrusion in 3D printing resulting in mechanically weak prints as well as over-extrusion causing excess use of material with little strength gain	Utilization of deep-learning-based computer vision system to correct under- and over-extrusion issues commonly found in 3D printing technology such as the fused deposition modeling (FDM)	The implementation of this model shows that the correction of under and over-extrusion errors when they happen thanks to the application of the developed system leads to a sixfold increase in print consistency while increasing print strength by up to 200%, reducing excess print material, and saving up to 40% material cost.
[19]	Use inexpensive webcams and a single multi-head deep convolutional neural network to augment any extrusion-based 3D printer with error detection, correction, and parameter discovery for new materials.	CAXTON system for autonomous data collection A network of eight FDM 3D printers were used for data collection. Creality CR-20 Pro printers were chosen due to their low cost, preinstalled bootloader and included Z probe.	Training a multi-head neural network using images automatically labelled by deviation from optimal printing parameters. The thus trained neural network, alongside a control loop, enables real-time detection and rapid correction of diverse errors that is effective across many different 2D and 3D printers.
[20]	Use of soft sensor that was 3D printed, development of an AI-powered 3D printing system that adapts to changes and movements of target surface, and use of hydrogel-based EIT sensor to monitor lung deformations.	Combination of "offline" and "online" machine learning vision-based tracking for estimating surface deformation.	3D printing system that estimates the motion and deformation of the target surface to adapt the toolpath in real time. Therefore improving: Precision, energy efficiency, versatility and reducing waste.
[21]	SLA process uses a digital micromirror device to project a set of mask images onto the resin surface to cure layers. After solidification of each layer, the building platform moves down at a predefined amount for the next layer.	To predicting, learning, and compensating 3-D shape deviations based on data, there was proposed shape deviation generator (SDG), a data-analytical framework to facilitate the learning and prediction of 3-D printing shape accuracy.	To predict, learn, and compensate for 3-D shape deviations using shape measurement data, it was proposed a shape deviation generator (SDG) under a novel convolution formulation to facilitate the learning and prediction of 3-D printing accuracy.

The identified research gap stems from the absence of studies integrating motion tracking technology with cameras for failure detection in 3D printing processes. This void in the existing literature underscores the need for our research, as it addresses this deficiency by proposing an innovative approach that combines the application of motion tracking dots on prints that will later be analyzed by the cameras to autonomously identifying and halt a print one's failure is detected. This scientific gap demonstrates the significance of our approach to failure detection in 3D printing, showcasing a unique contribution to the field.

III. Proposed Solution for AI based Failure Detection in Sustainable 3D Printing

A. Conceptual Model of Agents, System Elements Interacting with Each Other

In the proposed solution, we foresee a dynamic interaction between many components, including 3D printer, motion tracking nodes, cameras, labeled and unlabeled data, and the monitoring system as illustrated in Fig. 1. The system's main key is the 3D printer, which leads to additive manufacturing processes. Motion tracking nodes are strategically positioned on the 3D printer and play a crucial role in tracking its motions and the behavior of the manufactured product. Cameras are used to record the real-time movement of these nodes and provide us with unlabeled data. Two types of data (labeled and non-labeled) undergo a comparison process, by projecting one date onto another and looking for deviations. The monitoring system oversees analyzing data, identifying abnormalities, and terminating a printing process in real time. During the 3D printing process, motion tracking notes continuously record the position and orientation of the printer's components during the 3D printing process. Collected information is then transmitted to the cameras, which check the movement and the position of these nodes. Therefore, the camera data is analyzed by monitoring system and compared to the predicted motion patterns. If deviations are detected, the system intervenes in real-time to stop the printing process, to assure print quality and waste reduction.

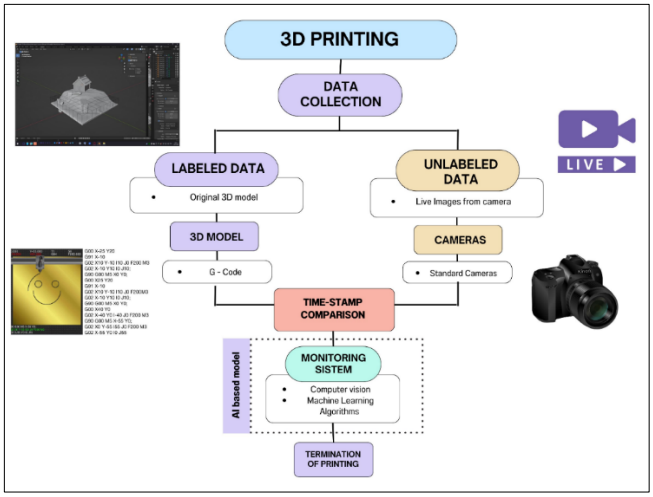


Fig. 1. The conceptual model

Our first hypothesis is that incorporating motion tracking technology into the 3D printing process can drastically improve the identification and mitigation of printing errors. Motion tracking technology, which has been used effectively in a variety of sectors, may be used to follow 3D printing in real time. It enables the detection of irregularities in the movement and location of printer components, which can serve as early warning signs of probable printing difficulties.

Our suggested solution tackles a research gap by bringing a fresh technique to 3D printing failure identification. Existing solutions frequently rely on post-print inspection or inadequate fault detection while printing. We can monitor the whole process in real time by including motion tracking nodes and cameras, recognizing problems as they occur. This novel technique makes a significant addition to the discipline by minimizing waste and improving 3D printing quality.

B. Framework for Detailed Solution

This research is based on the ontological premise that reality can be observed and measured. We think an epistemological stance that prioritizes empirical evidence and data-driven conclusions. The axiological stance cares for sustainability, efficiency, and the reduction of waste in 3D printing processes. The approach to theoretical development is essentially inductive. We will collect unlabeled data from real-world 3D printing processes, compare it with labeled data, and draw conclusions based on our findings. Additionally, we will apply a logical approach to expand on current ideas in 3D printing, motion tracking, and error detection.

The strategy is incorporating motion tracking technology into 3D printing. Data gathering with the help of motion tracking nodes and cameras, data processing using computer vision and machine learning algorithms, and the construction of real-time monitoring system are all part of the methodology. The experiment design will incorporate probability sampling methods to ensure a representative and unbiased selection of participants or data points, with the plan to print 10 test models for comprehensive evaluation.

The fusion of labeled and non-labeled data in the context of 3D printing failure detection involves a scientifically rigorous process to harness the strengths of both datasets. Our Labeled data, comprising the original 3D model, the G-code, provides explicit information about the expected finished product. Non-labeled data, acquired from cameras during the printing process, captures a broader range of real-world scenarios.

To facilitate the merging of labeled and non-labeled data detection, we propose a systematic approach. Initially, a machine learning model is trained using a labeled dataset that includes successful printing as well as varied failure occurrences. The model is then applied to unlabeled data in order to detect probable failure points.

The temporal synchronization of our labeled data, represented by the G-code, and our unlabeled data, captured by the cameras, is a pivotal aspect facilitating their seamless integration. Both datasets are equipped with time stamps, allowing for precise alignment of events during the 3D printing process. The time stamps act as temporal markers, ensuring a correspondence between the specific instructions encoded in the G-code and the corresponding visual information recorded by the cameras. This temporal alignment not only establishes a coherent chronological sequence but also enables the creation of a unified temporal framework for our data.

Active learning chooses samples with low model confidence for manual annotation, which enriches the labeled dataset. This merged dataset is utilized for model retraining, which improves the model's capacity to detect faults. Iterative self-training uses high-confidence predictions on leftover unlabeled data to extend the classified dataset. This cycling process improves the performance of the model, resulting in a continual learning loop for robust 3D printing failure verification. Regular assessment ensures that the model generalizes successfully and adjusts to new data patterns.

IV. Evaluation Scenarios

A. The Experiment Type for Your Research Project

Building upon the innovative methodology presented in [22-24], our proposed approach extends the capabilities of AI-based Computer Vision for failure detection in 3D printing. While the original work focused on identifying stringing defects during the printing process, our advancement involves the integration of labeled and non-labeled data using active learning and self-training. By incorporating time-stamped G-code information and camera feed data into our model, we enhance its capacity to detect a broader spectrum of failures in real-time.

This expansion not only allows for the identification of stringing but also facilitates the recognition of various other printing defects. Through a robust validation strategy, we aim to demonstrate improvements in both precision and recall, showcasing the efficacy of our approach in comparison to the existing model. Moreover, our proposed framework opens avenues for real-time adjustments to the printing process, enabling not only the termination of flawed prints but also corrective actions for parameters related to identified defects.

To achieve our goal of refining 3D printing failure detection, our first step will be to collect a diverse dataset comprising labeled data (G-code) and non-labeled data obtained from cameras monitoring the 3D printing process. Ensuring that both datasets possess time stamps for synchronization. After that we will preprocess the labeled data by extracting relevant features from the 3D models, for our non-labeled data, we will employ computer vision techniques to process the camera feed, extracting meaningful visual information.

Once we have our data, we will implement an active learning approach to strategically select informative instances from the non-labeled dataset for manual annotation. This iterative process optimizes the model's performance with minimal labeling effort. Moving forward we will apply self-training to iteratively improve the model's understanding by incorporating confidently predicted but initially non-labeled instances into the training process. This step enhances the model's adaptability to diverse printing scenarios. Using this information we will train a machine learning model, such as a Deep Convolutional Neural Network, using the merged dataset. We will utilize the labeled data for supervised learning and the

active learning and self-training iterations to enhance the model's performance.

In order to ensure the implementation of feature alignment techniques for compatibility between the labeled and non-labeled data we will map the feature spaces of both datasets to a common domain, facilitating seamless integration. To validate our results, we will split the dataset into training, validation, and test sets. Employing cross-validation techniques and select appropriate evaluation metrics (precision, recall, etc.) for a comprehensive assessment of the model's performance considering real-world testing to evaluate the model's robustness.

B. Validation Approach: Employing Data Augmentation and Model Evaluation in Real-Time Defect Detection

After gathering training data for the stringing defect, we implemented various Data Augmentation techniques to increase the number of training instances. Data Augmentation comprises a set of techniques applied to existing datasets to generate new synthetic data with meaningful information real, and its application was deemed necessary in this context. Specifically, for each image in the initial training set, we employed scaling to reduce the size (image resize), horizontal flipping, random cropping, 90-degree rotation, and random brightness adjustments. By applying these five data augmentation techniques to each of the 500 original training images, we expanded the dataset to a total of 2500 images.

There is a diverse selection of cutting-edge algorithms currently accessible, exhibiting differences in training speed, accuracy, and testing speed across benchmark datasets. In our case, we consider, use of cameras the necessity of striking a balance between achieving high accuracy and swift detection, given that the intended use of the model involves deployment in a real-time environment. In this investigation, the chosen model was the Single Shot Detector [22].

The Single Shot Detector running on 300 × 300 input (SSD-300), published in 2016, achieved a mean Average Precision (mAP) of 74.3% on benchmark Dataset VOC-2007 at 59 frames per second (FPS) and a mean Average Precision (mAP) of 41.2% at an Intersection over Union (IoU) of 0.5 on benchmark Dataset of Common Objects in Context (COCO test-dev2015).

V. Conclusion

In conclusion, our current study addresses critical challenges in integrating motion tracking technologies for real-time monitoring in 3D printing. While providing valuable insights, there remain opportunities for further exploration and improvement in this field.

It is imperative to thoroughly evaluate the research questions posed in this study, identifying areas that warrant additional investigation. These areas include:

- Extending the Scope of Motion Tracking: Explore the feasibility of incorporating advanced motion tracking technologies beyond the nodes-camera configuration to enhance precision and robustness in real-time monitoring.
- The Influence on Printing Speed and Efficiency: Examine how motion tracking affects the overall speed and efficiency of the 3D printing process, investigating potential impacts on printing speed and total production efficiency.
- Material Considerations: Investigate the interaction of various 3D printing materials with motion tracking systems, accounting for material differences and printing settings to assess the suggested system's flexibility across a wider range of materials.
- Human Interaction and Intervention: Examine the role of human contact and intervention when combined with real-time monitoring, considering scenarios where human operators may need to act based on monitoring system feedback, with implications for scalability in industrial settings.
- Economic and Environmental Consequences: Conduct a comprehensive examination of the economic and environmental consequences of incorporating motion tracking for real-time monitoring, comparing cost-effectiveness and sustainability to standard mistake correction approaches.
- User Acceptance and Experience: Investigate the user experiences and adoption of motion tracking-based real-time monitoring by 3D printer operators. Explore factors influencing user adoption, potential challenges, and ways to enhance user acceptability.

In summary, our study delves into the realm of 3D printing, which has evolved significantly since its inception in 1971, with fused filament fabrication (FFF) being the most widely used type today. Despite its affordability, FFF faces printing problems that threaten its commercial viability and environmental sustainability. Our proposed real-time monitoring system, utilizing motion tracking nodes and cameras, aims to address this gap by dynamically identifying and mitigating printing issues as they occur.

This research envisions a robust system for real-time intervention, ensuring print quality and waste reduction by integrating labeled G-code data with non-labeled data from cameras. The proposed framework employs an inductive theoretical development strategy, combining machine learning, computer vision, and motion tracking technologies. The experimental design includes active learning and self-training iterations, demonstrating the model's improved flaw detection over time. Our findings align with the transformative potential of 3D printing, emphasizing the need for individualized and flexible production. In conclusion, our suggested system pioneers real-time monitoring in 3D printing, utilizing motion tracking, computer vision, and machine learning. As we move into the implementation phase, we anticipate that our study will contribute to the evolution of 3D printing technologies, enhancing efficiency, sustainability, and adaptability to diverse production demands.

REFERENCES

[1] Petsiuk, A. L. & Pearce, J. M. Open source computer vision-based layer-wise 3D printing analysis. Addit. Manuf. 36, 101473 (2020).

[2] Jin, Z., Zhang, Z. & Gu, G. X. Autonomous in-situ correction of fused deposition modeling printers using computer vision and deep learning. Manuf. Lett. 22, 11–15 (2019).

[3] Rezaei, M. R., Houshmand, M. & Fatahi Valilai, O. An autonomous intelligent framework for optimal orientation detection in 3D printing. Int. J. Comput. Integr. Manuf. (2023) doi:10.1080/0951192X.2022.2162587.

[4] Liu, J. et al. A survey on location and motion tracking technologies, methodologies and applications in precision sports. Expert Syst. Appl. 229, 120492 (2023).

[5] Valizadeh, S., Fatahi Valilai, O., Houshmand, M. & Vasegh, Z. A novel digital dentistry platform based on cloud manufacturing paradigm. Int. J. Comput. Integr. Manuf. 32, 1024–1042 (2019).

[6] Valizadeh, S., Fatahi Valilai, O. & Houshmand, M. Allocation and scheduling of digital dentistry services in a dental cloud manufacturing system. Int. J. Comput. Integr. Manuf. 35, 645–661 (2022).

[7] Kucukdeger, E. & Johnson, B. N. Closed-loop controlled conformal 3D printing on moving objects via tool-localized object position sensing. J. Manuf. Process. 89, 39–49 (2023).

[8] Rezaei, M. reza, Houshmand, M. & Fatahi Valilai, O. An autonomous framework for interpretation of 3D objects geometric data using 2D images for application in additive manufacturing. PeerJ Comput. Sci. 7, e629 (2021).

[9] Arbabian, M. E. & Wagner, M. R. The impact of 3D printing on manufacturer–retailer supply chains. Eur. J. Oper. Res. 285, 538–552 (2020).

[10] Yuan, Y. Research Status and Development Trend of 3D Printing Technology. IOP Conf. Ser. Mater. Sci. Eng. 711, 012014 (2020).

[11] Thakar, C. M. et al. 3d Printing: Basic principles and applications. Mater. Today Proc. 51, 842–849 (2022).

[12] Customized Production Through 3D Printing in Cloud Manufacturing - 1st Edition. https://shop.elsevier.com/books/customized-production-through-3d-printing-in-cloud-manufacturing/zhang/978-0-12-823501-0.

[13] Shahrubudin, N., Lee, T. C. & Ramlan, R. An Overview on 3D Printing Technology: Technological, Materials, and Applications. Procedia Manuf. 35, 1286–1296 (2019).

[14] AbouelNour, Y. & Gupta, N. Assisted defect detection by in-process monitoring of additive manufacturing using optical imaging and infrared thermography. Addit. Manuf. 67, 103483 (2023).

[15] Muñiz Castro, B. et al. Machine learning predicts 3D printing performance of over 900 drug delivery systems. J. Controlled Release 337, 530–545 (2021).

[16] Li, C. et al. Fusing convolutional generative adversarial encoders for 3D printer fault detection with only normal condition signals. Mech. Syst. Signal Process. 147, 107108 (2021).

[17] Ghasemi, A. & Naser, M. Z. Tailoring 3D printed concrete through explainable artificial intelligence. Structures 56, 104850 (2023).

[18] Taskazan, B., Miller, J., Inyang-Udoh, U., Camps, O. & Sznaier, M. Domain Adaptation Based Fault Detection in Label Imbalanced Cyberphysical Systems. in 2019 IEEE Conference on Control Technology and Applications (CCTA) 142–147 (2019). doi:10.1109/CCTA.2019.8920608.

[19] Lee, W. et al. Adaptive 3D Printing for In Situ Adjustment of Mechanical Properties. Adv. Intell. Syst. 5, 2200229 (2023).

[20] Vaid, H., Shubham, P., Sharma, P. & Vajpeyi, R. Prediction of Printing Parameters for Minimal Dimensional Variation in 3D Printed Parts by Using Artificial Intelligence. in 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN) 350–354 (2021). doi:10.1109/SPIN52536.2021.9566052.

[21] Paraskevoudis, K., Karayannis, P. & Koumoulos, E. P. Real-Time 3D Printing Remote Defect Detection (Stringing) with Computer Vision and Artificial Intelligence. Processes 8, 1464 (2020).

[22] Brion, D. A. J. & Pattinson, S. W. Generalisable 3D printing error detection and correction via multi-head neural networks. Nat. Commun. 13, 4654 (2022).

[23] Zhu, Z., Park, H. S. & McAlpine, M. C. 3D printed deformable sensors. Sci. Adv. 6, eaba5575 (2020).

[24] Huang, Q., Wang, Y., Lyu, M. & Lin, W. Shape Deviation Generator—A Convolution Framework for Learning and Predicting 3-D Printing Shape Accuracy. IEEE Trans. Autom. Sci. Eng. 17, 1486–1500 (2020).

Developing a Blockchain Based Supply Chain CO_2 Footprint Tracking Framework Enabled by IoT

Mohammad Yaser Mofatteh[1], Roshanak Davallou[2], Chaida Ndahiro Ishimwe[3],
Swaresh Suresh Divekar[4], Omid Fatahi Valilai[5]*
School of Business, Social & Decision Science, Constructor University Bremen, Campus Ring 1, 28759, Bremen, Germany

***Abstract*—In various industries, the convergence of the Internet of Things (IoT) and blockchain technologies has left an indelible mark on the pursuit of decarbonization. These innovations have seamlessly integrated into diverse fields, from manufacturing to logistics, offering sustainable solutions that enhance operational efficiency, transparency, and accountability. The interplay between IoT and blockchain has particularly contributed to the reduction of carbon footprints, fostering environmentally responsible practices. As industries embrace these technologies, the decentralized and transparent nature of blockchain ensures traceability in supply chains, while IoT devices facilitate real-time data monitoring. Together, they create a powerful synergy that not only streamlines processes but also drives a collective commitment to reducing environmental impact, marking a paradigm shift towards greener and more sustainable industries. Within this landscape, this research offers a comprehensive exploration of the transformative potential of blockchain in supply chain management, emphasizing its intricate connection with IoT and carbon footprint reduction. The conceptual model presented delineates the seamless integration of these elements, providing a nuanced understanding of how blockchain can revolutionize transparency and sustainability. Through practical examples and a layered diagram, it showcases the tangible benefits of this integration, highlighting its capacity to enhance data integrity and transparency in real-world supply chain scenarios. The research stands as a testament to the instrumental role that blockchain can play in fostering environmentally responsible practices within supply chains, laying the groundwork for a more sustainable future.**

Keywords*—Blockchain; IoT; supply chain; carbon footprint; sustainability

I. Introduction

The management and reduction of CO_2 emissions present a multifaceted challenge for organizations, as they grapple with diverse emission sources and the urgency of climate change mitigation [1]. Tracking CO_2 emissions across the entirety of the supply chain introduces complexity due to the global nature of supply networks, encompassing numerous suppliers and intricate logistics networks [2]. Ensuring accuracy, consistency and transparent emissions data throughout the supply chain is a formidable task, complicated by variations in reporting methodologies, data availability, and the reluctance of some stakeholders to disclose emissions information. Also, how IoT devices can communicate directly through autonomous data will be challenging because the reports and data are usually ready to use [1].

In an era of escalating environmental concerns, the logistics industry stands at the forefront of innovation, striving to reduce its carbon footprint and enhance its overall sustainability [3]. The convergence of emerging technologies, such as blockchain and the Internet of Things (IoT), holds immense promise in transforming logistics operations towards greener and more efficient practices [4], [5]. This deductive research proposal outlines an extensive investigation into the economic and environmental benefits of utilizing blockchain for managing and optimizing green logistics processes, as well as the pivotal role that blockchain plays in verifying the authenticity and accuracy of sustainability claims and certifications within the logistics industry [6]. The Internet of Things (IoT) also emerged as a powerful tool that promises not only to revolutionize the way logistics operations are managed but also to substantively contribute to global sustainability goals [7], [8].

This research is anchored by three central research questions. First, aims to elucidate the economic and environmental benefits brought about by the incorporation of blockchain technology into green logistics processes. Second, it delves into the role of blockchain in validating sustainability claims and certifications within the logistics industry, while concurrently examining the autonomous data generation capabilities of IoT devices. Lastly, it explores the mechanisms and technologies that facilitate autonomous data generation by IoT devices to support field devices in logistics operations. These questions provide the guiding framework for our research inquiry.

The initial theory posits that the integration of blockchain technology, in conjunction with the utilization of the Internet of Things (IoT), has the potential to revolutionize the green logistics industry by offering solutions to key challenges. This theory is founded on the premise that blockchain can enhance transparency, reduce costs, and improve the accuracy of sustainability claims within the logistics sector. Additionally, IoT technologies, when harnessed effectively, can autonomously generate and share data, thereby optimizing field devices' operation and decision-making processes. It is hypothesized that the synergy between blockchain and IoT, it is hypothesized, can lead to substantial economic and environmental benefits for organizations engaged in green logistics operations. Through rigorous research and analysis, the paper seeks to validate and refine this initial theory, offering valuable insights into sustainable business practices in the context of logistics and environmental responsibility.

To provide a clear roadmap for the subsequent sections of this research paper [9], it commences with an extensive review of existing literature, establishing the theoretical foundations of the study. Following that, the research methodology will be delineated, elucidating the processes of data collection, analysis, and interpretation. Mixed methodologies such as rigorous literature reviews, incisive case studies, targeted surveys, and collaborative research will be used. Subsequently, the research presents findings, addressing each research question individually [10]. Finally, it will be concluded with a comprehensive discussion, drawing connections between the research outcomes and their broader implications for green logistics, blockchain technology, and IoT applications.

II. Literature Review

A. Green Logistics

Green logistics, often referred to as environmentally sustainable logistics or eco-logistics, in the field of logistics and supply chain management, green logistics has become a crucial concept due to the rising challenges of environmental sustainability [10]. In the past, the logistics industry mainly concentrated on making operations more efficient and economical, sometimes neglecting environmental concerns. Nevertheless, as environmental problems like climate change and resource depletion became more important, there was a significant change in how logistics was perceived and carried out [11], [12].

Green logistics is a major change in how we do logistics. It's all about making sure that by moving things around, it doesn't harm the environment. This means doing things like minimizing pollution, using less energy, and not making as much waste. At the same time, it still be wanted to use resources wisely and make money [13]. This change shows that now logistics should be considered as a way to help the planet and make money, and there is a need to think about both things as the world changes.

B. Blockchain Technology

Blockchain Technology can be defined as a distributed and digital ledger technology that transparently records and verifies every step in the supply chain [14]. It offers a secure and immutable record-keeping system that enables real-time tracking of environmentally friendly practices, facilitates transparency in sustainability reporting, and helps organizations make informed decisions to minimize their carbon footprint and enhance ecological responsibility throughout their logistics operations [15]. Blockchain technology offers innovative solutions for reducing carbon footprints and fostering sustainability within supply chains. By providing end-to-end traceability and transparency, blockchain enables companies to pinpoint inefficiencies, optimize transportation routes, and select environmentally responsible suppliers [16]. Smart contracts automate eco-friendly actions, such as choosing energy-efficient shipping options [17]. The immutability of blockchain data ensures the accuracy of emissions records, preventing fraudulent sustainability claims [18]. Decentralization reduces intermediaries, cutting emissions associated with unnecessary transportation [19]. These blockchain-driven solutions empower organizations to simultaneously reduce their carbon footprint and bolster supply chain efficiency and sustainability, aligning business goals with environmental responsibility [20].

C. Literature Gap Analysis

1) Blockchain based sustainable models/applications. The pursuit of sustainable models underpinned by blockchain technology presents a transformative opportunity, particularly when coupled with the integration of IoT autonomous data. Within this context, one promising model is the "Smart contracts and emission token system" [17]. A blockchain-based sustainable model that combines smart contracts and emission tokens can significantly enhance sustainability efforts by providing a transparent, automated, and secure framework for tracking, verifying, and incentivizing emissions reductions [21].

In this innovative model for combating climate change and promoting sustainability, emissions are transformed into digital tokens, symbolizing the reduction or removal of greenhouse gases from the environment [22]. These tokens are generated based on scientifically validated emission reduction projects or actions. Crucial to this system are smart contracts, which lay out the terms and conditions of these emission reduction initiatives, specifying emission reduction goals, the methods for verification, and the issuance of emission tokens upon achieving these goals [23].

By combining smart contracts with emission tokens on a blockchain-based platform, this sustainable model enhances transparency, security, and automation in emissions tracking and reduction efforts [24]. It also encourages organizations and individuals to actively participate in sustainable practices and carbon offsetting while contributing to the global fight against climate change [25].

2) IoT and autonomous data integration with blockchain. A major change in the landscape of sustainable logistics is being brought about by the convergence of IoT, autonomous data generation, and blockchain integration [26]. The diverse potential of these technologies in solving operational and environmental concerns within supply chains has recently been highlighted.

Smart sensors used in the IoT for real-time monitoring have become popular because of their capacity to increase efficiency and transparency [27]. These strategically positioned sensors provide ongoing information on environmental conditions, safeguarding the integrity of goods and materials. Vehicle performance sensors, for instance, improve fleet management and lower carbon emissions. IoT devices, equipped with sensors and connected to networks, autonomously collect and transmit data in real-time [28]. These devices are capable of capturing environmental variables, vehicle performance metrics, and supply chain data, among others. Their ability to transmit data without manual intervention is a foundational component in achieving autonomous data communication.

Decentralized networks like blockchain play a vital role in enabling IoT devices to communicate autonomously. These networks, often peer-to-peer in nature, facilitate direct data

exchange between devices [29], [30]. This eliminates the need for intermediaries and central authorities, streamlining communication while enhancing security and efficiency also, IoT device-generated autonomous data enables data-driven decision-making for logistics operations via direct communication, the optimization of routes and predictive maintenance.

These developments are enhanced by blockchain technology, which offers a transparent and unchangeable ledger for storing data and autonomous data communication generated by IoT devices. It serves as a distributed ledger where data is securely recorded, and time stamped. This ensures data integrity and authenticity, crucial in green logistics, particularly when validating sustainability claims or ensuring compliance with environmental regulations [26]. By guaranteeing the validity of environmental data, this integration strengthens the plausibility of sustainability claims and avoids "greenwashing" [31]. The result of literature study based on the defined themes is discussed in Table I.

TABLE I. LITERATURE STUDY BASED ON THE MAIN THEMES

Research Studies	Green logistics	Blockchain Technology application	IoT and autonomous interactions
Liu and Ma [28]	New energy-saving transportation methods.		IoT application to guide warehousing operations to improve work efficiency.
Rosado Da Cruz et al. [21]	Tracing carbon footprints.	Blockchain technology for monitoring, tracking and tracing the carbon footprints of products. Smart contracts implementation.	
Tan et al. [15]	Green logistics is based on blockchain by integrating IoT and big data in logistics industry.	Applications of blockchain in logistics operations. Transforming traditional operations into blockchain based operations.	
Feroz et al. [18]	Minimizing the CO_2 emission and managing energy.	Blockchain provides prominent levels of privacy, security and traceability.	
Chelh and Ababou [32]		Potential to revolutionize supply chain management by reducing risks like fraud and data breaches while creating a resilient in green supply chain.	
Cruz [26]		Convergence of IoT-Block chain models to enhance data privacy.	
Nizetic et al. [33]			Using IoT devices to simplify process and to have efficiency, sustainability, and quality.
Ahmad [34]		Combining IoT devices and blockchain technology on Ethereum to enable secure, decentralized data exchange and automation with smart contracts in a P2P network, brings transparency.	
This research	Measuring, evaluation, and controlling of CO_2 footprint through supply chain material operations.	Application of blockchain for transparency of material data flow.	Autonomous data communication and interoperability, data privacy and security, connectivity and standardization, data driven decision making with IoT analytics.

The existing literature highlights the potential of integrating IoT autonomous data collection with blockchain technology to create sustainable and transparent supply chains. However, a significant research gap exists in the development and validation of comprehensive models and applications that can be readily adopted by organizations. These models should effectively integrate IoT autonomous data with blockchain, ensuring data accuracy, transparency, and trust throughout the supply chain [32]. To address this gap, the research will focus on bridging the divide between theory and practical application, providing adaptable, user-friendly frameworks for organizations looking to leverage IoT and blockchain for sustainable, environmentally friendly and transparent supply chains [33].

III. PROPOSED SOLUTION

The development of the frameworks should be the primary focus of this research to ensure that the promise of IoT, autonomous data, and blockchain integration is fully realized in sustainable logistics, with substantial advantages for both businesses and the environment. Fig. 1 demonstrates the conceptual model of such framework.

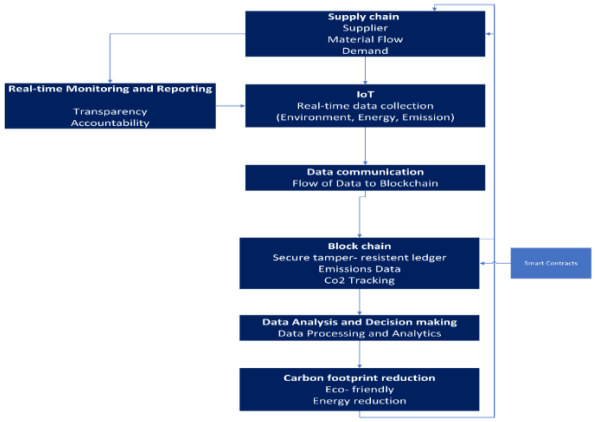


Fig. 1. Illustration of IoT, blockchain conceptual model for reducing overall CO_2 in supply chain.

A. Supply Chain Layer

The Supply Chain Layer consists of suppliers, demand and material flow. It ensures efficient sourcing, monitors material movement and forecasts demand, to enhance overall supply chain performance and sustainability [18]. This layer plays a

pivotal role in leveraging insights for informed decision-making [34].

B. Real-Time Monitoring and Reporting Layer

Throughout the process, real-time monitoring and reporting are facilitated by this layer. Stakeholders have access to emissions data, ensuring transparency and accountability [35].

C. IoT Layer

In the next layer, IoT devices are deployed throughout the supply chain. These devices autonomously collect real-time data with RFID or NFC communication methods on environmental conditions, energy consumption, and emissions [3].

D. Data Communication Layer

The data collected by IoT devices is transmitted to the Data Communication Layer. This layer manages the flow of data from the IoT devices to the blockchain [33].

E. Blockchain Layer

The data will be received and processed within the blockchain layer. The blockchain acts as a secure and tamper-resistant ledger, recording and time stamping all emissions data. Smart contracts, self-executing pieces of code on the blockchain, enforce predefined rules and automate processes. In supply chain and sustainability, smart contracts can execute actions based on the received data, such as verifying emissions compliance or triggering alerts for specific thresholds. On the other side, Emission tokens quantify a certain amount of carbon footprint emissions associated with a specific activity, process, or entity [36]. These tokens are part of the broader concept of carbon credits or carbon offset mechanisms, where organizations seek to compensate for their carbon emissions by investing in projects that reduce or capture an equivalent amount of greenhouse gases elsewhere. Emission tokens often leverage blockchain technology, providing a transparent and immutable ledger to track the generation, transfer, and retirement of these tokens. Blockchain enhances the integrity and traceability of carbon-related data [37].

F. Data Analysis and Decision Layer

The processed data is then sent to the Data Analysis and Decision Layer. Here, data analytics tools are used to gain insights into emissions patterns and identify opportunities for reducing carbon footprints [38].

G. Carbon Footprint Reduction Layer

The insights and recommendations from the Data Analysis and Decision Layer inform the actions taken in the Carbon Footprint Reduction Layer. These actions can include selecting eco-friendly transportation methods, reducing energy consumption, or adjusting sourcing practices [39].

IV. Evaluation

A. Validation Model

The validation of this research asserts the robustness and credibility of the proposed model, affirming the veracity of our claims. The focus will be mainly on the order and transport transactions as can be seen in Fig. 2.

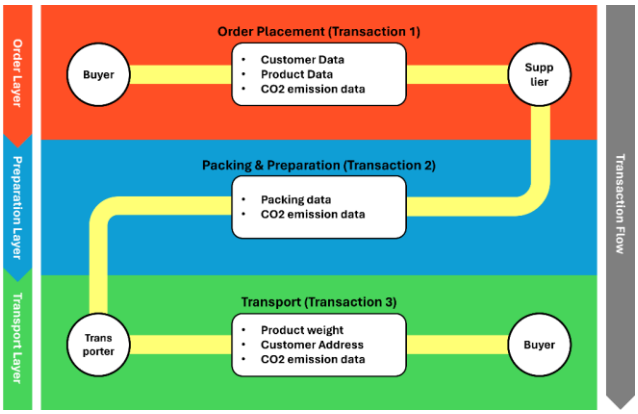


Fig. 2. Transaction and data flow of validation model for the proposed solution.

1) Order placement (Transaction 1). The initial transaction occurs when the buyer (which could be a distributor, retailer, or end consumer) places an order with the supplier. This transaction involves the transfer of order details, quantities, and specifications.

2) Packing and preparation (Transaction 2). The supplier initiates a transaction related to packing and preparing the goods for shipment. This may include generating packing lists, labeling, and preparing documentation.

3) Transport (Transaction 3). When the goods are ready for shipment, the supplier initiates a transaction to notify the transportation entity or logistics provider. This transaction includes information about the shipment, such as weight, dimensions, and the destination.

This approach has several advantages:

- Interoperability. Utilizing a general blockchain provides the advantage of fostering compatibility with diverse blockchain networks. This approach facilitates interoperability with other systems and applications that conform to prevalent blockchain standards.
- Ease of Integration. General blockchains typically feature well-established APIs and tools for network interaction. This simplifies the integration process, enabling concentration on the particular data intended to be stored in the blockchain, rather than navigating the intricacies of blockchain development.
- Security and Immutability. Leveraging an existing blockchain network assures the inherent security and immutability characteristics within blockchain technology. The decentralized and tamper-resistant nature of the blockchain ensures the integrity of the embedded information.
- Cost Efficiency. Opting for an existing blockchain network proves to be a more cost-effective choice. Creating a dedicated blockchain network can be resource-intensive, making the utilization of an established network a pragmatic alternative.

Overall, this approach not only simplifies the development process but also leverages the strengths of blockchain

technology without the need for extensive customization. It allows us to focus on the specific data and use case while benefiting from the broader blockchain ecosystem.

B. Results

The study yields a basic yet instructive blockchain implementation tailored to a supply chain context. Example transactions, such as order placement, packing, and transport, are utilized to showcase how data is structured within blocks and how blocks are sequentially linked. The creation of a genesis block and the utilization of cryptographic hashing underscore the foundational principles of blockchain technology.

The code serves as a steppingstone for comprehending more intricate blockchain frameworks and their real-world applications and highlights the importance of data integrity, decentralization, transaction transparency and key attributes of blockchain technology that contribute to its potential in optimizing supply chain management.

V. Conclusion

In this comprehensive exploration of blockchain integration in supply chain management, it is navigated through fundamental concepts and practical implementations, offering a nuanced understanding of the transformative potential of blockchain technology. The discussion commenced with the delineation of a conceptual model intricately weaving together the realms of the Internet of Things (IoT), blockchain, and carbon footprint reduction. The layers, ranging from IoT data collection to carbon footprint reduction strategies, provided a holistic perspective on how blockchain could revolutionize transparency, traceability, and sustainability within supply chain operations.

The ensuing creation and visualization of a layered diagram reinforced the intricate connections between IoT-enabled data acquisition, secure transmission through blockchain, data analysis for informed decision-making, and the imperative action layer focused on carbon footprint reduction. This conceptual model stands as a testament to the nuanced orchestration required to imbue supply chains with efficiency, accountability, and environmental responsibility. The focus on a supply chain scenario with example transactions elucidated the potential for blockchain to enhance data integrity and transparency in real-world applications. Encryption of blockchain data stands out as a crucial avenue for future research studies. Exploring robust encryption methods to restrict outsider access ensures that sensitive supply chain information remains secure and confidential. This avenue addresses growing concerns regarding data privacy and confidentiality within blockchain networks.

References

[1] D. Trihinas, L. Thamsen, J. Beilharz, and M. Symeonides, ‘Towards Energy Consumption and Carbon Footprint Testing for AI-driven IoT Services’, 2022 IEEE Int. Conf. Cloud Eng. IC2E, pp. 29–35, Sep. 2022, doi: 10.1109/IC2E55432.2022.00011.

[2] H. Thirunavukkarasu, K. Rajendran, K. Dinesh, and S. Gunasekaran, ‘Application of Blockchain and IoT in Cap and Trade of Carbon Emissions’, 2023 Int. Conf. Recent Adv. Electr. Electron. Ubiquitous Commun. Comput. Intell. RAEEUCCI, pp. 1–5, Apr. 2023, doi: 10.1109/RAEEUCCI57140.2023.10134302.

[3] V. Varriale, A. Cammarano, F. Michelino, and M. Caputo, ‘Sustainable Supply Chains with Blockchain, IoT and RFID: A Simulation on Order Management’, Sustainability, vol. 13, no. 11, p. 6372, Jun. 2021, doi: 10.3390/su13116372.

[4] S. Singh, P. K. Sharma, B. Yoon, M. Shojafar, G. H. Cho, and I.-H. Ra, ‘Convergence of blockchain and artificial intelligence in IoT network for the sustainable smart city’, Sustain. Cities Soc., vol. 63, p. 102364, Dec. 2020, doi: 10.1016/j.scs.2020.102364.

[5] S.-A. Radmanesh, A. Haji, and O. Fatahi Valilai, ‘Blockchain-Based Architecture for a Sustainable Supply Chain in Cloud Architecture’, Sustainability, vol. 15, no. 11, Art. no. 11, Jan. 2023, doi: 10.3390/su15119072.

[6] A. Arshad, F. Shahzad, I. Ur Rehman, and B. S. Sergi, ‘A systematic literature review of blockchain technology and environmental sustainability: Status quo and future research’, Int. Rev. Econ. Finance, vol. 88, pp. 1602–1622, Nov. 2023, doi: 10.1016/j.iref.2023.07.044.

[7] S.-A. Radmanesh, A. Haji, and O. F. Valilai, ‘Blockchain-based cloud manufacturing platforms: A novel idea for service composition in XaaS paradigm’, PeerJ Comput. Sci., vol. 7, p. e743, Dec. 2021, doi: 10.7717/peerj-cs.743.

[8] ETH Zurich, Switzerland et al., ‘Blockchain for the IoT: Privacy-Preserving Protection of Sensor Data’, J. Assoc. Inf. Syst., pp. 1272–1307, 2019, doi: 10.17705/1jais.00567.

[9] P. Basu, P. Deb, and A. Singh, ‘Blockchain and the carbon credit ecosystem: sustainable management of the supply chain’, J. Bus. Strategy, Jan. 2023, doi: 10.1108/JBS-09-2022-0157.

[10] B. Q. Tan, F. Wang, J. Liu, K. Kang, and F. Costa, ‘A Blockchain-Based Framework for Green Logistics in Supply Chains’, Sustainability, vol. 12, no. 11, p. 4656, Jun. 2020, doi: 10.3390/su12114656.

[11] C. Liu and T. Ma, ‘Green logistics management and supply chain system construction based on internet of things technology’, Sustain. Comput. Inform. Syst., vol. 35, p. 100773, Sep. 2022, doi: 10.1016/j.suscom.2022.100773.

[12] C. R. Lobo, H. Wicaksono, and O. F. Valilai, ‘Implementation of Blockchain Technology to Enhance Last Mile Delivery Models with Sustainability Perspectives’, 10th IFAC Conf. Manuf. Model. Manag. Control MIM 2022, vol. 55, no. 10, pp. 3304–3309, Jan. 2022, doi: 10.1016/j.ifacol.2022.10.123.

[13] C. Qian, Y. Gao, and L. Chen, ‘Green Supply Chain Circular Economy Evaluation System Based on Industrial Internet of Things and Blockchain Technology under ESG Concept’, Processes, vol. 11, no. 7, p. 1999, Jul. 2023, doi: 10.3390/pr11071999.

[14] M. S. Asif and H. Gill, ‘Blockchain Technology and Green Supply Chain Management (GSCM) – Improving Environmental and Energy Performance in Multi-echelon Supply Chains’, IOP Conf. Ser. Earth Environ. Sci., vol. 952, p. 012006, Jan. 2022, doi: 10.1088/1755-1315/952/1/012006.

[15] B. Q. Tan, F. Wang, J. Liu, K. Kang, and F. Costa, ‘A Blockchain-Based Framework for Green Logistics in Supply Chains’, Sustainability, vol. 12, p. 4656, Jun. 2020, doi: 10.3390/su12114656.

[16] A. Arshad, F. Shahzad, I. Ur Rehman, and B. S. Sergi, ‘A systematic literature review of blockchain technology and environmental sustainability: Status quo and future research’, Int. Rev. Econ. Finance, vol. 88, pp. 1602–1622, Nov. 2023, doi: 10.1016/j.iref.2023.07.044.

[17] S. Aggarwal and N. Kumar, ‘Blockchain 2.0: Smart contracts’, in Advances in Computers, vol. 121, Elsevier, 2021, pp. 301–322. doi: 10.1016/bs.adcom.2020.08.015.

[18] A. K. Feroz, H. Zo, and A. Chiravuri, ‘Digital Transformation and Environmental Sustainability: A Review and Research Agenda’, Sustainability, vol. 13, no. 3, p. 1530, Feb. 2021, doi: 10.3390/su13031530.

[19] M. Singh, G. Baranwal, and A. K. Tripathi, ‘Decentralized group decision making using blockchain’, J. Supercomput., vol. 79, no. 17, pp. 20141–20178, Nov. 2023, doi: 10.1007/s11227-023-05426-6.

[20] S. Ganesan, H. Wicaksono, and O. Fatahi Valilai, ‘Enhancing Vendor Managed Inventory with the Application of Blockchain Technology’, in

Advances in System-Integrated Intelligence, M. Valle, D. Lehmhus, C. Gianoglio, E. Ragusa, L. Seminara, S. Bosse, A. Ibrahim, and K.-D. Thoben, Eds., Cham: Springer International Publishing, 2023, pp. 262–275.

[21] A. Rosado Da Cruz, F. Santos, P. Mendes, and E. Cruz, 'Blockchain-based Traceability of Carbon Footprint: A Solidity Smart Contract for Ethereum':, in Proceedings of the 22nd International Conference on Enterprise Information Systems, Prague, Czech Republic: SCITEPRESS - Science and Technology Publications, 2020, pp. 258–268. doi: 10.5220/0009412602580268.

[22] L. Swinkels, 'Trading Carbon Credit Tokens on the Blockchain', SSRN Electron. J., 2023, doi: 10.2139/ssrn.4378871.

[23] N. Zhao, Z. Sheng, and H. Yan, 'Emission trading innovation mechanism based on blockchain', Chin. J. Popul. Resour. Environ., vol. 19, no. 4, pp. 369–376, Dec. 2021, doi: 10.1016/j.cjpre.2022.01.010.

[24] G. Palaiokrassas et al., 'Combining Blockchains, Smart Contracts, and Complex Sensors Management Platform for Hyper-Connected SmartCities: An IoT Data Marketplace Use Case', Computers, vol. 10, no. 10, p. 133, Oct. 2021, doi: 10.3390/computers10100133.

[25] F. Zhao and W. K. (Victor) Chan, 'When Is Blockchain Worth It? A Case Study of Carbon Trading', Energies, vol. 13, no. 8, p. 1980, Apr. 2020, doi: 10.3390/en13081980.

[26] A. Cruz, 'Convergence between Blockchain and the Internet of Things', Int. J. Technol. Innov. Manag. IJTIM, vol. 1, no. 1, pp. 34–53, Sep. 2021, doi: 10.54489/ijtim.v1i1.11.

[27] S. R. Niya, S. S. Jha, T. Bocek, and B. Stiller, 'Design and implementation of an automated and decentralized pollution monitoring system with blockchains, smart contracts, and LoRaWAN', NOMS 2018 - 2018 IEEEIFIP Netw. Oper. Manag. Symp., pp. 1–4, Apr. 2018, doi: 10.1109/NOMS.2018.8406329.

[28] C. Liu and T. Ma, 'Green logistics management and supply chain system construction based on internet of things technology', Sustain. Comput. Inform. Syst., vol. 35, p. 100773, Sep. 2022, doi: 10.1016/j.suscom.2022.100773.

[29] A. Rejeb, J. G. Keogh, and H. Treiblmaier, 'Leveraging the Internet of Things and Blockchain Technology in Supply Chain Management', Future Internet, vol. 11, no. 7, p. 161, Jul. 2019, doi: 10.3390/fi11070161.

[30] A. Rejeb et al., 'Unleashing the power of internet of things and blockchain: A comprehensive analysis and future directions', Internet Things Cyber-Phys. Syst., vol. 4, pp. 1–18, 2024, doi: 10.1016/j.iotcps.2023.06.003.

[31] A. Inês, A. Diniz, and A. C. Moreira, 'A review of greenwashing and supply chain management: Challenges ahead', Clean. Environ. Syst., vol. 11, p. 100136, Dec. 2023, doi: 10.1016/j.cesys.2023.100136.

[32] P. Dutta, R. Chavhan, P. Gowtham, and A. Singh, 'The individual and integrated impact of Blockchain and IoT on sustainable supply chains:a systematic review', Supply Chain Forum Int. J., vol. 24, no. 1, pp. 103–126, Jan. 2023, doi: 10.1080/16258312.2022.2082851.

[33] K.-H. Liu, S.-F. Chang, W.-H. Huang, and I.-C. Lu, 'The Framework of the Integration of Carbon Footprint and Blockchain: Using Blockchain as a Carbon Emission Management Tool', in Technologies and Eco-innovation towards Sustainability I: Eco Design of Products and Services, A. H. Hu, M. Matsumoto, T. C. Kuo, and S. Smith, Eds., Singapore: Springer, 2019, pp. 15–22. doi: 10.1007/978-981-13-1181-9_2.

[34] K. Govindan, D. Kannan, T. B. Jørgensen, and T. S. Nielsen, 'Supply Chain 4.0 performance measurement: A systematic literature review, framework development, and empirical evidence', Transp. Res. Part E Logist. Transp. Rev., vol. 164, p. 102725, Aug. 2022, doi: 10.1016/j.tre.2022.102725.

[35] 'Addressing the Challenges of Real-Time Data Sharing in IoT - RTInsights'. Accessed: Oct. 16, 2023. [Online]. Available: https://www.rtinsights.com/addressing-the-challenges-of-real-time-data-sharing-in-iot/

[36] 'Blockchain for Climate Change and Carbon Monitoring: A Game-Changer in Environmental Sustainability | by Solidity Academy | Medium'. Accessed: Nov. 18, 2023. [Online]. Available: https://medium.com/@solidity101/blockchain-for-climate-change-and-carbon-monitoring-a-game-changer-in-environmental-68c2153c2223

[37] 'Token Emission and Its Importance - LCX'. Accessed: Nov. 18, 2023. [Online]. Available: https://www.lcx.com/token-emission-and-its-importance/

[38] I. H. Sarker, 'Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective', SN Comput. Sci., vol. 2, no. 5, p. 377, Sep. 2021, doi: 10.1007/s42979-021-00765-8.

[39] S. Srisruthi, N. Swarna, G. M. S. Ros, and E. Elizabeth, 'Sustainable agriculture using eco-friendly and energy efficient sensor technology', in 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India: IEEE, May 2016, pp. 1442–1446. doi: 10.1109/RTEICT.2016.7808070.

Triggered Screen Restriction Framework: Transforming Gamified Physical Interventions

Majed Hariri[1], Richard Stone[2], Ulrike Genschel[3]
Department HCI, Iowa State University, Ames, USA[1]
Department Industrial Engineering, Iowa State University, Ames, USA[2]
Department Statistics, Iowa State University, Ames, USA[3]

***Abstract*—This study examines the effectiveness of the Triggered Screen Restriction (TSR) framework, a novel technique to promote exercise that combines negative reinforcement with adaptive gamification elements. The study examined the TSR framework's impact on physical activity levels, addictive nature, health indicators, psychological factors, and app usability compared to a control group. A mixed experimental design was employed, with 30 participants randomly assigned to either an experimental group using a custom iOS app with the TSR framework or a control group using a similar app without TSR features. Results revealed that the TSR group demonstrated significantly higher physical activity levels ($p < .05$). The TSR framework resulted in significant increases in app usage frequency ($p < .001$). Health indicators showed a significant improvement in balance and stability through the single-leg stance test ($p < .05$), while other health metrics, including maximum jumping jacks completed in one minute, post-exercise heart rate, and body composition, exhibited no significant changes. Analysis of psychological factors revealed a significant increase in perceived competence in the TSR group ($p < .05$), with no significant changes observed in autonomy or relatedness. The TSR intervention demonstrated significantly better usability metrics, including ease of use, system reliability, and perceived usefulness, compared to the control condition (all $p < .001$). The study contributes to the expanding adoption of gamified physical interventions, showcasing the TSR framework as an effective technique for addressing physical inactivity. Future research should explore long-term effectiveness, diverse populations, and integration with wearable devices to further validate and refine the TSR approach in addressing physical inactivity.**

***Keywords*—Gamification; physical activity; negative reinforcement; triggered screen restriction framework; TSR framework; gamified physical intervention**

I. Introduction

Gamification incorporates elements from games into non-gaming settings. Its goal is to boost user engagement, motivation, and participation [1]. Increasingly, gamification is being used in physical activity interventions. Gamified interventions use game-like elements such as badges and rankings to encourage people to be more physically active [2]. Points, for instance, are awarded for task completion, symbolizing achievements. Leaderboards compare users' performances, promoting competition by ranking their achievements. Badges, earned through specific actions, visually represent achievements. Progress is marked by levels, showing advancement, and providing feedback. Challenges are set goals crafted to increase user engagement and active participation. By transforming physical activity into a game-like experience, the gamified intervention aims to increase participation and address the global decline in physical activity and its consequent rise in sedentary lifestyles [3], [4].

Gamification draws inspiration from motivational psychology and strategically employs intrinsic and extrinsic motivators to enhance user engagement in non-game contexts. By embedding some game elements, the intervention stimulates participation through the inherent desire to explore, learn, and enjoy, which indicates intrinsic motivation [5], [6]. Intrinsic motivation promotes self-regulation, as individuals are driven by personal interest instead of external rewards [7], [8]. Conversely, extrinsic motivation makes people do things for external rewards. These rewards can be money or praise from others [9]. The self-determination theory (SDT) helps us understand motivation better. SDT says that individuals have three main needs. These are the need to make choices, to feel capable, and to connect with others. Meeting these needs is key to increasing motivation and overall happiness [10], [11]. Autonomy emphasizes control over one's actions, competence pertains to the ability to navigate challenges effectively, and relatedness indicates a sense of connection with others [12], [13], [14]. Through the lens of SDT, gamification elements such as challenges and points serve as mechanisms to fulfill these innate psychological needs, thereby enhancing motivation and participation in physical activities [5], [15], [16].

A. Gamification Frameworks

The Octalysis framework identifies eight primary motivators for individuals. These motivators encourage individuals to take part in activities. The framework presents these motivators visually as an octagon [17], [18], [19]. It categorizes these motivators into white hat (positive motivations) and black hat (negative motivations) gamification, emphasizing the importance of balancing both for effective gamification.

The Mechanics, Dynamics, and Aesthetics framework encapsulates various game tactics and examination methods. It links game development, analysis, and technological investigation [20], [21], [22]. The Mechanics, Dynamics, and Aesthetics framework divides games into three main parts. Mechanics are the core elements and algorithms of the game. Dynamics refer to how the mechanics respond to player inputs in real-time. Aesthetics involve the emotional reactions of players.

The Sustainable Gamification Design framework aspires to create gamified systems with long-lasting benefits. It focuses

on three main areas. These are user activity, environmental influence, and social commitment [23], [24], [25], [26]. The Sustainable Gamification Design framework contains four key stages: Discover, Reframe, Envision, and Create, emphasizing moral and human-centered layout guides throughout the process.

The FRAGGLE framework is an agile method. The FRAGGLE framework aims to improve learning experiences through gamification. This framework aligns gamified actions with educational objectives, content, and assessment measures [27], [28], [29]. The framework has four main phases. These phases are Declaration, Creation, Execution, and Learning. The framework focuses on making gamified activities that are both engaging and educationally effective.

The growing interest in gamification for promoting physical activity presents a promising opportunity. A novel framework can be molded to encourage exercise. The novel framework can address specific needs and motivations. It can do so by using negative reinforcement. By incorporating gamification elements, this innovative approach has the prospect of boosting the acceptance of healthy habits significantly.

B. The TSR Framework

The TSR framework offers a new strategy for implementing gamification elements in exercises. It addresses the limitations of traditional methods [30]. This framework combines adaptive gamification strategies, machine learning models, and computer vision. It creates an intriguing and personalized experience to promote physical exercise. The TSR framework stands out due to its unique features. These include negative reinforcement, adaptive gamified components, and real-time action verification that respects privacy [31].

The primary technique of the TSR framework leverages individuals' Fear of Missing Out (FOMO) and employs negative reinforcement to enhance motivation. It restricts entry to social media apps based on exercise objectives. This technique harnesses people's urge to remain connected and educated. Physical activity becomes essential to access restricted venues [32]. The TSR approach also includes customized message triggers. It uses a model of computer vision to exercise recognition. The adaptive incentive system adjusts the difficulty level based on each user's established routine [31].

The TSR framework builds on existing research on gamified physical interventions. It adds new elements to overcome common limitations. We need to explore several areas to understand the TSR framework's contributions fully. These include the current landscape of gamified physical interventions, how negative reinforcement affects behavior change, and the addictive nature of electronic devices. The next section reviews related work in these areas. This process provides context for the TSR framework's unique approach and its potential to promote physical activity.

II. Related Work

A. Gamified Physical Interventions

A gamified intervention incorporating points, progress bars, challenges, and leaderboards was investigated for its effects on daily step counts. The results indicated a significant increase in steps taken by participants during the intervention period [33], [34], [35].

Researchers conducted a controlled trial study using a gamified physical intervention on overweight adults. The intervention used a fitness tracker and a smartphone app. It gave game-like feedback, helped set goals, and offered social support. The study found that during the intervention, participants took more steps. However, this increase did not last after the intervention ended [36], [37], [38].

Researchers investigated a gamification-based intervention designed to encourage walking among obese adults. The results indicated that participants who engaged with the gamified elements of the intervention increased their walking frequency [39], [40], [41].

The efficacy of gamification in enhancing physical activity was examined using a commercially available activity tracker coupled with a custom-designed mobile app. The gamified intervention, which incorporated points, leaderboards, and social interaction features, significantly increased step counts [42], [43].

A gamified intervention utilizing challenges and self-monitoring tools was designed to encourage physical activity in adults. The study identified a significant positive effect of the gamified intervention in increasing participants' daily step counts [44], [45].

The use of a gamified social media intervention delivered through WeChat was examined to promote physical activity. The intervention utilized various game mechanics, including competition, points, and social interaction, to encourage walking. The findings demonstrated that the intervention effectively increased participants' daily step counts [46].

The long-term effects of a gamified intervention on physical activity maintenance were investigated. The intervention involved points, progress bars, competition, and financial rewards, leading to a significant increase in step counts. Notably, the positive effects persisted for a two-year follow-up period, emphasizing the potential for long-term behavioral change [47], [48].

Researchers investigated a gamified intervention to increase walking in overweight and obese adults. They used the Way to Health platform. Even though the program had game-like features, it did not lead to participants walking more [49].

Researchers assessed the impact of an app with game-like features to encourage adults to be more active. The app used points, progress bars, challenges, rankings, and rewards. The study found that participants using the app became more physically active [50].

A gamified app employing points, progress bars, challenges, and social interaction elements was created to encourage physical activity. The results showed a notable rise in physical activity among users who utilized the gamified app [51].

Researchers studied the impact of the game-like features to promote exercise. The intervention used points, progress bars, and challenges. The study showed that participants who used gamified physical intervention apps exercised more [51].

A study examined the long-term impact of a gamified cycling exercise. Despite using points and progress bars, the study did not show significant long-term significance on cycling activity [52].

The related studies analyzed the prospect of gamified physical interventions. Many of the mentioned gamified physical interventions overly rely on positive reinforcement, which led to a decline in activity levels post-intervention.

B. Negative Reinforcement

Negative reinforcement makes a behavior more likely to happen again by removing unwanted provocation. Understanding how negative reinforcement affects an individual's motivation to initiate new behaviors or tasks is crucial for education, therapy, and behavior modification applications.

Reinforcement theory implies that behavior can be modified through reinforcement. Negative reinforcement is one of the four primary strategies in reinforcement theory, alongside positive reinforcement, punishment, and extinction. It is effective in increasing desired behaviors by eliminating negative conditions, but it must be carefully managed to avoid unintended results [53], [54]. Combining positive and negative reinforcement is very effective in boosting task compliance and decreasing escape-maintained damaging behavior [55], [56]. The frequency of negative reinforcement can moderately affect the persistence in completing tasks. This implies that regular use can sustain desired behaviors over time [57].

Negative reinforcement can lead to distinct emotional responses, such as increased positive affect and reduced fatigue, which may influence motivation and feedback processing in organizational settings [58]. The negative reinforcement paradigm can be effectively used in classroom settings to increase desirable behaviors and decrease undesirable ones, suggesting its potential applicability in training and development programs [59].

Teachers and therapists can use negative reinforcement effectively to increase treatment integrity and compliance with interventions. A study has demonstrated that combining performance feedback with negative reinforcement improves implementation accuracy and reduces off-task behavior among students [60]. Interventions based on negative reinforcement can be customized to fit individual needs. This customization boosts their effectiveness in behavior modification programs [61].

C. Addictive Nature of Electronic Devices

The addictive nature of electronic devices, particularly smartphones, is a growing concern. Smartphone features like infinite scroll and autoplay facilitate the perception of reinforcement rewards and promote habit formation, which together drive smartphone addiction [62]. Users often spend more time than planned on applications with infinite scrolling due to the psychological effects of the feature, leading to increased dependency on these applications [63].

FOMO greatly contributes to the addictive nature of electronic devices. FOMO is linked with the destructive effects of social media on everyday life and job performance, with various social media use disorders acting as intermediaries [64]. FOMO causes people to frequently check their phones and can predict no-mobile-phone phobia [65]. No-mobile-phone is a contemporary psychological condition marked by the fear of not having a mobile phone [66]. It also mediates between psychological needs and phubbing, strengthening the habit of constantly checking smartphones [67].

Perceived enjoyment, emotional gain, and social pressure greatly contribute to smartphone addiction. Enjoyment from using smartphones can create habitual usage, which then leads to smartphone addiction [68]. Emotional gain, which includes positive emotions and relief from negative psychological states, is a strong predictor of smartphone addiction across generations [69].

Social media interactions offer positive reinforcement (for instance, likes and comments) and negative punishment (such as social comparison). The positive and negative factors contribute to social media dependence [70]. Individuals with social anxiety and loneliness are more prone to smartphone addiction, especially when their primary use is accessing social networking sites [71].

The extensive use of smartphones for social interactions and process-related activities, such as browsing and gaming, significantly contributes to habitual and addictive smartphone behavior [72], [73]. Younger generations exhibit higher levels of smartphone addiction compared to older generations, with emotional gain and social environment pressure being strong predictors of this behavior among younger users [74], [69].

III. Objective

The study aims to thoroughly assess the TSR framework, a novel gamification framework to promote physical activity. It will compare the TSR framework with a control group that undergoes a gamified physical intervention without using the TSR framework. The study seeks to determine whether the TSR framework leads to significant improvements in physical activity levels, health metrics, and psychological aspects. Additionally, it aims to evaluate the framework's impact on app usability and its potentially addictive nature. The study examines changes in psychological aspects such as autonomy, competence, and relatedness among participants using surveys conducted before and after the intervention. Furthermore, the study evaluates participants' experiences with the gamified physical intervention, focusing on the ease of use, system reliability, and usefulness of the TSR framework in promoting physical activity.

IV. Methodology

The study employed a mixed experimental design with between-subjects and within-subjects components to assess the effectiveness of the TSR framework in enhancing physical activity and health outcomes. Participants were randomly assigned to either an experimental group using a custom iOS app with the TSR framework or a control group.

A. Participants

Following Institutional Review Board approval, 30 participants (19 to 38 years old, M = 29.56, SD = 5.25)

were recruited from Iowa State University students and residents (see Table I). Participants were healthy adults without pre-existing medical conditions that could hinder physical activity. Recruitment was conducted through campus flyers and snowball sampling. Participants were informed that their involvement was voluntary and that they could withdraw at any time without consequences.

TABLE I. PARTICIPANT DEMOGRAPHICS

Characteristic	Category	Percentage
Education	Graduate degree	53%
	Bachelor's degree	30%
	High school diploma	17%
Employment Status	Employed full-time	60%
	Unemployed	3%
	Students	37%
Used Gamified Apps Before	Yes	53%
	No	47%
Competitive Features	Yes	77%
	No	23%

B. Procedure

1) Orientation and group assignment: All participants attended an orientation session. They learned about the study's goals and steps. Then, they were split into two groups randomly. The experimental group used a custom-made iOS app with the TSR framework. The control group used the same app but without the TSR framework.

2) Intervention period: The study duration was four weeks. During the study, both groups were asked to do jumping jacks daily using the given app. The app logged the number of repetitions for each participant.

3) Data collection: Participants attended two in-person sessions at the ATHENA Lab, located at 0066 Black Engineering, ISU, for measurement of stamina and physique indicators (maximum jumping jacks in one minute, post-exercise heart rate, single-leg stance test, and body composition). Psychological factors and app usability were assessed through surveys on a secure device in the lab.

4) Data management: Data was de-identified at the point of collection, with all measures associated with assigned study ID codes rather than participant names or contact information.

5) Debriefing: All participants were fully debriefed about the true nature of the research and the importance of having a control group. In the debriefing session, participants were informed of the experiment's aim to compare the effectiveness of a gamified physical intervention with and without the TSR in boosting physical levels and improving health outcomes. The research team explained that the deception was needed to avoid potential biases and ensure the study results were valid. Participants were encouraged to ask questions and share any concerns about the deception or the study in general. The research team addressed these questions and concerns openly and honestly.

C. Apparatuses

The study utilized several apparatuses to measure participants' health metrics and facilitate the experimental procedures. Body composition, including muscle mass and fat percentage, was assessed using the eufy by Anker, Smart Scale P1. This smart scale provides accessible body composition measurements, allowing for tracking changes in muscle mass and fat percentage throughout the visits. The study used the FACEIL Pulse Oximeter Fingertip, a non-invasive device that measures heart rate and oxygen saturation levels to measure the heart rate. The pulse oximeter measured participants' heart rate immediately after the one-minute jumping jacks exercise and at one, two, three, and four minutes post-exercise to assess cardiovascular recovery. An iPhone 14 Pro was used to precisely time the one-minute jumping jacks exercise and the post-exercise heart rate measurements, ensuring consistency and accuracy in the data collection process.

D. Measurements

The study assessed various variables to measure the significance of the TSR framework:

1) Addictive nature: Number of times the app was opened by each user.

2) Physical activity metrics: Number of jumping jack repetitions.

3) Health indicators: Maximum jumping jacks in one minute, post-exercise heart rate, single-leg stance test, and body composition.

4) Psychological factors: Autonomy, competence, and relatedness.

5) App usability measures: Ease of use, system reliability, and perceived usefulness.

The surveys were administered in person at the ATHENA Lab. Participants completed the surveys on a secure device provided in the ATHENA lab. These surveys were designed to collect participant responses securely and confidentially. All the surveys had the same Likert scale. Participants fill out 5-point questionnaires varying from "Disagree" to "Agree". The surveys were adjusted to ensure no personally identifiable information (like names or email addresses) was collected with the responses.

The scales measuring psychological factors (autonomy, competence, and relatedness) were adapted from previously validated instruments, the Basic Psychological Needs in Exercise Scale (BPNES) [75]. The app usability scales were also based on established usability questionnaires, the Technology Acceptance Model (TAM), [76] [77].

In prior studies, these validated scales showed strong psychometric indicators, such as high internal consistency reliability and construct validity. By employing these validated scales, the study ensured that the psychological factors and app usability measures were reliable and valid, allowing for meaningful interpretation of the results.

The physical activity metrics and health indicators were considered objective measures. Their validity was assumed based on the accuracy of the app's logging system and the standardized protocols used for the jumping jacks test, single-leg stance test, and body composition assessment.

V. Result

All data were analyzed using JASP version 0.18.3. We employed both independent samples t-tests and repeated measures ANOVA to assess differences in our outcome measures. Independent samples t-tests were used to evaluate between-group differences, while repeated measures ANOVA was utilized to analyze changes over time and interactions between time and app mode (TSR framework vs. control). Tukey's HSD test was used for post-hoc comparisons to evaluate our hypotheses. As the results were qualitatively consistent across both parametric and non-parametric methods, we present the t-test and ANOVA results due to their familiarity to most readers and their established robustness against violations of normality and homogeneity of variance, particularly when sample sizes are equal across treatment groups, as is the case in our study [78], [79].

A. Addictive Nature

H1. The TSR framework leads to a significant increase in app usage frequency compared to a gamified physical intervention without the TSR framework.

An independent samples t-test was conducted to evaluate the impact of the TSR framework on app usage frequency. The analysis revealed a significant difference in app usage between the experimental group and the control group [$t(28) = -5.552$, $p < .001$]. Participants in the experimental group ($M = 3.112$, $SD = 1.642$) opened the app significantly more often than those in the control group ($M = 0.626$, $SD = 0.558$). These results suggest that the TSR framework significantly impacts the addictive nature of the intervention, which resulted in higher usage frequency (see Fig. 1).

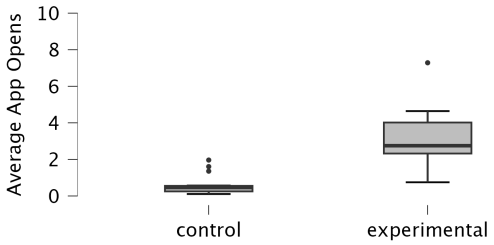


Fig. 1. Average app usage frequency by group.

B. Physical Activity Levels

H2. The TSR framework leads to a significant improvement in physical activity levels among participants compared to a control group without the TSR intervention.

An independent samples t-test was conducted to evaluate the impact of the TSR framework on physical activity levels, measured by the average number of repetitions performed. The analysis revealed a significant difference in repetitions between the experimental group and the control group [$t(28) = -2.346$, $p < .05$]. Participants in the experimental group ($M = 12.721$, $SD = 18.825$) performed significantly more repetitions on average than those in the control group ($M = 1.267$, $SD = 1.754$). These results suggest that the TSR framework significantly impacts physical activity levels (see Fig. 2).

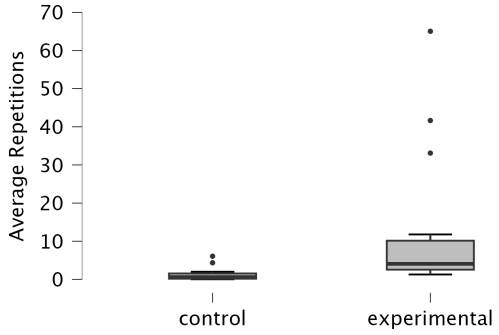


Fig. 2. Average exercise repetitions by group.

C. Health Indicators

H3. The TSR framework leads to a significant improvement in health metrics among participants compared to a control group experiencing a gamified physical intervention without the TSR framework.

Repeated measures ANOVA was performed to assess the effect of the intervention on all health metrics: maximum jumping jacks completed in one minute, post-exercise heart rate, single-leg stance test, and body composition. The analysis revealed a significant interaction between time and app mode in the within-subject effect for the Single-leg stance test [$F(1, 28) = 5.572$, $p < .05$]. The between-subjects analysis also showed a significant main effect of app mode for this test [$F(1, 28) = 7.509$, $p < .05$]. Post hoc comparisons using Tukey's HSD test revealed a significant difference between the before and after intervention scores for the experimental group (Mean Difference = -50.167, $p < .05$), while the control group showed no significant change. These results suggest that the TSR framework significantly improved balance and stability as measured by the Single-leg stance test. No significant changes were observed in the other health metrics.

D. Psychological Factors

H4. The TSR framework leads to a significant improvement in perceived psychological factors in physical activities compared to a gamified physical intervention without the TSR framework.

Repeated measures ANOVA was performed to assess the effect of the intervention on all psychological factors: autonomy, competence, and relatedness. The analysis revealed a significant main effect of time [$F(1, 28) = 7.303$, $p < .05$] and a significant interaction between time and app mode [$F(1, 28) = 4.534$, $p < .05$] for competence. Post hoc comparisons using Tukey's HSD test showed a significant increase in competence scores for the experimental group from before to after the intervention (Mean Difference = -0.656, $p < .05$), while the control group showed no significant change. These results suggest that the TSR framework significantly improved perceived competence in physical activities (see Fig. 3). No significant changes were observed for autonomy or relatedness.

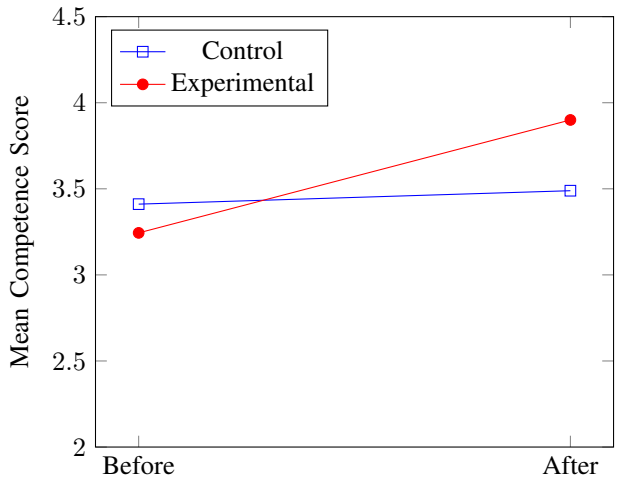


Fig. 3. Changes in competence scores before and after intervention.

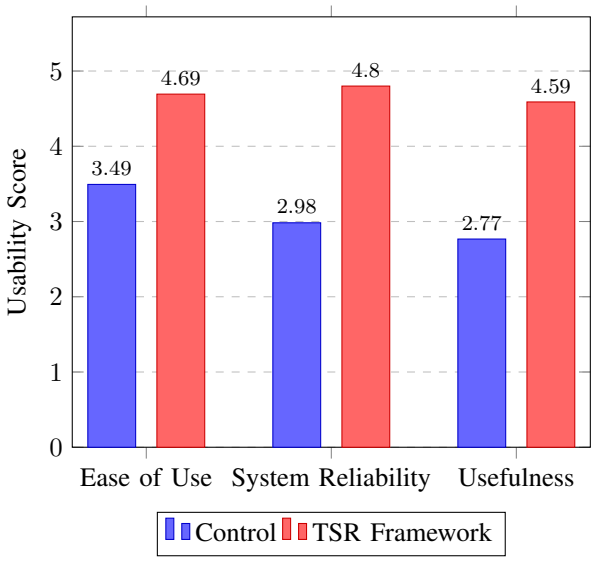


Fig. 4. Comparison of intervention usability factors between control and TSR framework groups.

E. App Usability

H5. The TSR framework leads to a significant improvement in perceived app usability compared to a gamified physical intervention without the TSR framework.

Independent samples t-tests were performed to assess the effect of the intervention on all intervention usability factors: ease of use, system reliability, and perceived usefulness. The analyses revealed significant differences between the experimental and control groups for all three factors. For perceived ease of use, a significant difference was found [t(28) = -4.372, p < .001], with the experimental group (M = 4.693, SD = 0.369) reporting higher scores than the control group (M = 3.493, SD = 0.997). Perceived system reliability also showed a significant difference [t(28) = -6.864, p < .001], with the experimental group (M = 4.800, SD = 0.330) scoring higher than the control group (M = 2.983, SD = 0.970). Similarly, perceived usefulness demonstrated a significant difference [t(28) = -5.042, p < .001], with the experimental group (M = 4.589, SD = 0.620) reporting higher scores than the control group (M = 2.767, SD = 1.255). The usability analysis shows the TSR framework intervention outperformed the control intervention in ease of use, reliability, and perceived usefulness. Participants reported it more user-friendly, reliable, and valuable for their physical activity goals, potentially boosting engagement, program commitment, and health outcomes (see Fig. 4).

VI. Discussion

This research offers important insights into the effectiveness of the TSR framework. The TSR approach to gamified physical interventions has shown promising results. The study's findings demonstrate notable improvements across several measures. However, the research also identifies specific areas that require additional investigation and refinement.

The significant increase in physical activity levels, as measured by jumping jack repetitions, in the experimental group compared to the control group aligns with previous research on gamified interventions for physical activity [33], [42], [44]. This result indicates that the TSR framework's unique approach effectively boosts users' engagement in physical exercise.

The analysis of app usage frequency revealed a significant increase in the number of times users opened the app in the experimental group compared to the control group. This finding supports the statement that through its innovative use of FoMO and negative reinforcement, the TSR framework might create an engaging and might lead to habit-forming intervention. The substantially higher usage frequency demonstrates the framework's success in leveraging addictive features to promote regular engagement with physical activities. By channeling the compelling nature of digital interactions into health-promoting behaviors, the TSR framework shows promise in transforming potentially problematic usage patterns into sustained, healthy habits.

Regarding health indicators, the significant improvement in the single-leg stance test for the experimental group is particularly noteworthy. The enhancement in balance and stability suggests that the TSR framework may offer benefits beyond just increasing activity levels, with potential implications for some physical functions. However, the lack of significant changes in other health metrics, including maximum jumping jacks completed in one minute, post-exercise heart rate, and body composition, indicates that the framework's impact on broader health outcomes may be limited, at least within the time frame of this study.

The analysis of psychological factors revealed interesting outcomes, particularly in terms of perceived competence. The significant increase in competence scores, especially evident in the experimental group, suggests that the TSR framework effectively boosts users' confidence in their ability to engage in physical activities. However, the lack of significant changes in autonomy and relatedness scores indicates that these aspects of psychological need satisfaction may require further refinement within the TSR framework.

The intervention usability measures strongly supported the

TSR framework's effectiveness in terms of user experience. Significant improvements were observed in ease of use, system reliability, and perceived usefulness compared to the control intervention. The findings highlight the importance of user experience in promoting engagement with gamified physical interventions and suggest that the TSR framework successfully addresses key aspects of technology acceptance [77].

The research findings should be interpreted with caution due to certain constraints in the study design. A key limitation was the restricted number of participants involved, which may have affected the statistical power of the analysis. Additionally, the shortness of the experimental time frame could have obscured some outcomes, particularly for health-related metrics that typically necessitate longer observation periods to manifest noticeable changes. Finally, the lack of social features and the use of machine-learning techniques may have resulted in a lack of improvement in some factors, especially autonomy and relatedness.

Nevertheless, the research offers compelling evidence for the potential of the TSR framework in promoting physical activity and improving specific health outcomes. The framework's ability to significantly increase user engagement, improve balance and stability, enhance perceived competence, and provide a superior user experience suggests that it offers a promising approach to addressing physical inactivity.

VII. Future Work

The promising results of this study on the TSR framework highlight several avenues for future research and development:

1) Long-term effectiveness: Future studies should investigate the TSR framework's ability to sustain behavior changes over extended periods. A longitudinal study lasting several months could provide valuable insights into whether the framework can maintain increased physical activity levels and the improved health indicator observed in this study. The long-term direction addresses a common challenge in health interventions and could help validate the long-term impact of the TSR approach.

2) Diverse populations: Future research should include a broader range of participants, including different age groups, fitness levels, and backgrounds. Testing the TSR framework with older adults, children, or people with specific health conditions would enhance the generalizability of the findings and potentially reveal how the framework's effectiveness varies across different demographics.

3) Varied physical activities: While the current study focused on jumping jacks, future studies should explore the framework's efficacy with a wider range of physical activities. Future studies could include aerobic exercises, strength training, or sports movements, providing insights into the versatility and broader applicability of the TSR framework across different types of physical activity.

4) Psychological factors refinement: Given the significant improvement in perceived competence but lack of change in autonomy and relatedness, future work should focus on enhancing the TSR framework to better support all aspects of psychological need satisfaction. This might involve incorporating more personalized goal-setting features or embracing social interaction elements to address autonomy and relatedness more effectively.

5) Health indicators assessment: The lack of significant changes in most health indicators, except for the single-leg stance test, suggests a need for more comprehensive and possibly longer-term health assessments. Future research could employ more sensitive measures of cardiovascular fitness, muscular strength, and body composition. Additionally, adjusting the framework's intensity and duration of physical activities might lead to more substantial health improvements.

6) Addictive nature analysis: While the increased app usage frequency demonstrates the TSR framework's effectiveness in engaging users, further research is needed to explore the long-term implications of this engagement. Studies could investigate how the framework's addictive features translate into sustained healthy behaviors over time and examine potential strategies to optimize engagement without introducing problematic usage patterns.

7) Wearable device integration: Future work could explore combining the TSR framework with popular wearable fitness devices. The integration with wearable devices could provide more accurate activity tracking and personalized feedback, potentially enhancing the framework's effectiveness and user experience even further.

VIII. Conclusion

The study provides strong evidence for the effectiveness of the TSR framework as an innovative approach to gamified physical interventions. The results demonstrate that the framework's unique features can significantly impact physical activity levels and certain health indicators. Key findings of the study include:

- A significant increase in physical activity levels in the experimental group, as measured by jumping jack repetitions, compared to the control group.
- Significantly higher app usage frequency in the TSR framework group, demonstrating the framework's success in leveraging addictive features to promote regular engagement with physical activities.
- Improved balance and stability in the TSR framework group, evidenced by significant improvements in the single-leg stance test.
- Increased perceived competence among participants using the TSR framework, while no significant changes were observed in autonomy and relatedness.
- Superior user experience with the TSR framework intervention, with significantly higher ratings in ease of use, system reliability, and perceived usefulness compared to the control intervention.

In conclusion, the TSR framework addresses several limitations of conventional gamified physical interventions by using negative reinforcement and incorporating adaptive gamification elements. The TSR approach shows promise in creating a more engaging, potentially habit-forming method for promoting physical activity. The framework offers an innovative and impactful approach to encouraging healthier,

more active lifestyles. The TSR framework has been demonstrated as an effective method for tackling physical inactivity, contributing to the broader implementation of gamified physical interventions.

ACKNOWLEDGMENT

Majed Hariri's study efforts were sponsored by the Islamic University of Madinah.

REFERENCES

[1] Deterding, Sebastian and Dixon, Dan and Khaled, Rilla and Nacke, and Lennart, "From game design elements to gamefulness: defining" gamification"," *Proceedings of the 15th International Academic MindTrek Conference Envisioning Future Media Environments*, p. 341, 2011.

[2] J. Hamari, J. Koivisto, and H. Sarsa, "Does gamification work? - A literature review of empirical studies on gamification," in *Proceedings of the Annual Hawaii International Conference on System Sciences*. IEEE Computer Society, 2014, pp. 3025–3034.

[3] R. Guthold, G. A. Stevens, L. M. Riley, and F. C. Bull, "Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1·6 million participants," *The Lancet Child and Adolescent Health*, vol. 4, no. 1, pp. 23–35, 1 2020.

[4] A. M. Bisantz, J. Pfautz, R. Stone, E. M. Roth, G. Thomas-Meyers, and A. Fouse, "Assessment of display attributes for displaying metainformation on maps," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 3. SAGE Publications Sage CA: Los Angeles, CA, 2006, pp. 289–293.

[5] F. Groh, "Gamification: State of the art definition and utilization," *Institute of Media Informatics Ulm University*, vol. 39, p. 31, 2012.

[6] E. S. Abdelall, R. T. Stone, L. Rajana Nedved, and C. Fales, "Bend the tool, not the wrist: An ergonomic investigation of medical suturing needle holder design and medical engineering solution," *Surgical Innovation*, vol. 28, no. 1, pp. 62–70, 2021.

[7] L. Legault and M. Inzlicht, "Self-Determination, Self-Regulation, and the Brain: Autonomy Improves Performance by Enhancing Neuroaffective Responsiveness to Self-Regulation Failure," *Journal of Personality and Social Psychology*, vol. 105, no. 1, pp. 123–138, 7 2013.

[8] E. S. Abdelall, Z. Eagle, T. Finseth, A. A. Mumani, Z. Wang, M. C. Dorneich, and R. T. Stone, "The interaction between physical and psychosocial stressors," *Frontiers in behavioral neuroscience*, vol. 14, p. 63, 2020.

[9] Irving B. Weiner and W. Edward Craighead., "SELF-DETERMINATION," *The Corsini Encyclopedia of Psychology*, 2009.

[10] R. M. Ryan and E. L. Deci, "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being Self-Determination Theory," Tech. Rep., 1985.

[11] A. Mumani, R. Stone, and Z. Wang, "The effect of scanning technology and upc placement on supermarket self-checkout," *Packaging Technology and Science*, vol. 31, no. 2, pp. 83–96, 2018.

[12] L. Legault, "Self-Determination Theory," in *Encyclopedia of Personality and Individual Differences*. Cham: Springer International Publishing, 2017, pp. 1–9. [Online]. Available: http://link.springer.com/10.1007/978-3-319-28099-8_1162-1

[13] M. Sailer, J. U. Hense, S. K. Mayr, and H. Mandl, "How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction," *Computers in Human Behavior*, vol. 69, pp. 371–380, 4 2017.

[14] E. L. Deci and R. M. Ryan, *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media, 2013.

[15] S. Baumann, R. Stone, E. Abdelall, V. Srikrishnan, T. Schnieders, C. Fales, and A. Mumani, "Implementing blockchain to enhance usability of patient-generated data," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 63, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2019, pp. 1344–1348.

[16] S. Baumann, R. T. Stone, U. Genschel, and F. Mgaedeh, "The picon methodology applied: Operator errors and preference tracking of a novel ubiquitous vital signs sensor and its user interface," *International Journal of Human–Computer Interaction*, vol. 40, no. 14, pp. 3782–3804, 2024.

[17] Y.-k. Chou, *Actionable gamification: Beyond points, badges, and leaderboards*. Packt Publishing Ltd, 2019.

[18] T. M. Schnieders and R. T. Stone, "Current work in the human-machine interface for ergonomic intervention with exoskeletons," *International Journal of Robotics Applications and Technologies (IJRAT)*, vol. 5, no. 1, pp. 1–19, 2017.

[19] T. M. Schnieders, A. A. Mumani, R. T. Stone, and B. Westby, "An analytic network process model for ranking exoskeleton evaluation criteria," *Theoretical Issues in Ergonomics Science*, pp. 1–11, 2024.

[20] R. Hunicke, M. LeBlanc, R. Zubek *et al.*, "Mda: A formal approach to game design and game research," in *Proceedings of the AAAI Workshop on Challenges in Game AI*, vol. 4, no. 1. San Jose, CA, 2004, p. 1722.

[21] R. T. Stone, S. Pujari, A. Mumani, C. Fales, and M. Ameen, "Cobot and robot risk assessment (carra) method: an automation level-based safety assessment tool to improve fluency in safe human cobot/robot interaction," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 65, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2021, pp. 737–741.

[22] R. T. Stone, F. Z. Mgaedeh, and A. N. Pulley, "Cognitive and physiological evaluation of virtual reality training in nursing," *Ergonomics*, pp. 1–13, 2024.

[23] M. Raftopoulos, "Towards gamification transparency: A conceptual framework for the development of responsible gamified enterprise systems," *Journal of Gaming & Virtual Worlds*, vol. 6, no. 2, pp. 159–178, 2014.

[24] R. T. Stone, A. M. Bisantz, J. Llinas, and V. Paquet, "Augmented multisensory interface design (amid): A human-centric approach to unisensory and multisensory augmented reality design," *Journal of Cognitive Engineering and Decision Making*, vol. 3, no. 4, pp. 362–388, 2009.

[25] T. Schnieders, B. Moeller, A. Mumani, and R. Stone, "Vastus and patellar protection with range of motion pad–advanced personal protective equipment for the lower body," 2018.

[26] R. T. Stone, "Human power generation design assessment: an evaluation of ergonomic risk, metabolic burden, and overall design efficiency," *International Journal of Human Factors and Ergonomics*, vol. 1, no. 3, pp. 282–297, 2012.

[27] A. Mora, P. Zaharias, C. González, and J. Arnedo-Moreno, "Fraggle: a framework for agile gamification of learning experiences," in *Games and Learning Alliance: 4th International Conference, GALA 2015, Rome, Italy, December 9-11, 2015, Revised Selected Papers 4*. Springer, 2016, pp. 530–539.

[28] Z. Wang, R. Stone, J. Kim, and S. Baumann, "Financial orientated heuristic evaluation for hand tools 2021 hfes international annual meeting," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 65, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2021, pp. 859–863.

[29] M. Y. Al-Daraghmeh and R. T. Stone, "A review of medical wearables: materials, power sources, sensors, and manufacturing aspects of human wearable technologies," *Journal of Medical Engineering & Technology*, vol. 47, no. 1, pp. 67–81, 2023.

[30] M. Hariri and R. Stone, "Gamification in physical activity: State-of-the-art," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 10, 2023. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2023.01410105

[31] ——, "Trigger screen restriction framework, ios use case towards building a gamified physical intervention." *International Journal of Advanced Computer Science & Applications*, vol. 15, no. 5, 2024.

[32] "Triggered screen restriction: Gamification framework," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 11, 2023. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2023.01411130

[33] R. Nuijten, P. Van Gorp, A. Khanshan, P. Le Blanc, A. Kemperman, P. van den Berg, and M. Simons, "Health promotion through monetary incentives: Evaluating the impact of different reinforcement schedules on engagement levels with a mhealth app," *Electronics (Switzerland)*, vol. 10, no. 23, 12 2021.

[34] F. V. A. Srikrishnan, R. T. Stone, and C. Xu, "An engineering approach to storage and access in high priority scenarios," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 63, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2019, pp. 1349–1353.

[35] R. T. Stone, "Actively guided practice enhances kinesthetic memory development," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 57, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2013, pp. 640–644.

[36] M. S. Patel, D. S. Small, J. D. Harrison, V. Hilbert, M. P. Fortunato, A. L. Oon, C. A. Rareshide, and K. G. Volpp, "Effect of Behaviorally Designed Gamification with Social Incentives on Lifestyle Modification among Adults with Uncontrolled Diabetes: A Randomized Clinical Trial," *JAMA Network Open*, 2021.

[37] R. Stone, M. Vasan, F. Mgaedeh, Z. Wang, and B. Westby, "Evaluation of latest computer workstation standards," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 66, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2022, pp. 853–857.

[38] R. Stone, J. Kim, C. Xu, F. Mgaedeh, C. Fales, and B. Westby, "Effects of semi-automatic pistol slide pull device on law-enforcement racking process," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 66, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2022, pp. 903–907.

[39] A. K. Agarwal, K. J. Waddell, D. S. Small, C. Evans, T. O. Harrington, R. Djaraher, A. L. Oon, and M. S. Patel, "Effect of Gamification with and without Financial Incentives to Increase Physical Activity among Veterans Classified as Having Obesity or Overweight: A Randomized Clinical Trial," *JAMA Network Open*, vol. 4, no. 7, 7 2021.

[40] C. W. Fales, R. T. Stone, D. Van Groningen, and B. Westby, "Tread patterns and the effect on basketball player slippage," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 64, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2020, pp. 1328–1332.

[41] B. Westby, R. Stone, C. Fales, and D. Bonner, "Law enforcement uniforms and public perception: An overview and pilot study," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 67, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2023, pp. 2305–2309.

[42] X. S. Chen, S. Changolkar, A. S. Navathe, K. A. Linn, G. Reh, G. Szwartz, D. Steier, S. Godby, M. Balachandran, J. D. Harrison, C. A. Rareshide, and M. S. Patel, "Association between behavioral phenotypes and response to a physical activity intervention using gamification and social incentives: Secondary analysis of the STEP up randomized clinical trial," *PLoS ONE*, vol. 15, no. 10 October, 10 2020.

[43] R. Stone, B. Westby, N. Jaskowiak, and C. Fales, "Forensic footwear: A retrospective of the development of the mantis shoe scanning system," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications Sage CA: Los Angeles, CA, 2024, p. 10711813241265340.

[44] L. M. Lier and C. Breuer, "The motivating power of gamification: Does the inclusion of game elements increase the effectiveness of worksite health promotion programs?" *International Journal of Workplace Health Management*, vol. 13, no. 1, pp. 1–15, 1 2020.

[45] B. Westby and R. T. Stone, "Safety and performance analysis of non-residential roofing materials: Empirical evaluation of standards, norms, and assumptions," *Norms, and Assumptions*.

[46] D. Mo, M. Xiang, M. Luo, Y. Dong, Y. Fang, S. Zhang, Z. Zhang, and H. Liang, "Using gamification and social incentives to increase physical activity and related social cognition among undergraduate students in shanghai, china," *International Journal of Environmental Research and Public Health*, vol. 16, no. 5, p. 858, 2019.

[47] M. A. Harris, "Maintenance of behaviour change following a community-wide gamification based physical activity intervention," *Preventive Medicine Reports*, vol. 13, pp. 37–40, 3 2019.

[48] A. Mumani and R. Stone, "State of the art of user packaging interaction (upi)," *Packaging Technology and Science*, vol. 31, no. 6, pp. 401–419, 2018.

[49] G. W. Kurtzman, S. C. Day, D. S. Small, M. Lynch, J. Zhu, W. Wang, C. A. Rareshide, and M. S. Patel, "Social Incentives and Gamification to Promote Weight Loss: The LOSE IT Randomized, Controlled Trial," *Journal of General Internal Medicine*, vol. 33, no. 10, pp. 1669–1675, 10 2018.

[50] R. Nuijten, P. Van Gorp, A. Khanshan, P. Le Blanc, P. van den Berg, A. Kemperman, and M. Simons, "Evaluating the Impact of Adaptive Personalized Goal Setting on Engagement Levels of Government Staff With a Gamified mHealth Tool: Results From a 2-Month Randomized Controlled Trial," *JMIR mHealth and uHealth*, vol. 10, no. 3, 3 2022.

[51] Z. Zhao, A. Arya, R. Orji, G. Chan *et al.*, "Effects of a personalized fitness recommender system using gamification and continuous player modeling: system design and long-term validation study," *JMIR serious games*, vol. 8, no. 4, p. e19968, 2020.

[52] J. Berg, A. I. Wang, S. Lydersen, and T. Moholdt, "Can gaming get you fit?" *Frontiers in Physiology*, vol. 11, p. 1017, 2020.

[53] L. Shiguo, "Explaining the existent rationality of matthew effect in scientific study in the view of reinforcement theory," *Science and Technology Management Research*, 2011.

[54] M. F. Villere and S. S. Hartman, "Reinforcement theory: A practical tool," *Leadership & Organization Development Journal*, vol. 12, pp. 27–31, 1991.

[55] K. J. Bouxsein, H. Roane, and T. Harper, "Evaluating the separate and combined effects of positive and negative reinforcement on task compliance." *Journal of applied behavior analysis*, vol. 44 1, pp. 175–9, 2011.

[56] C. C. Piazza, W. W. Fisher, G. P. Hanley, M. L. Remick, S. A. Contrucci, and T. L. Aitken, "The use of positive and negative reinforcement in the treatment of escape-maintained destructive behavior." *Journal of applied behavior analysis*, vol. 30 2, pp. 279–97; quiz 297–8, 1997.

[57] P. W. Romani, J. Ringdahl, D. Wacker, N. H. Lustig, K. M. Vinquist, J. Northup, A. M. Kane, and D. P. Carrion, "Relations between rate of negative reinforcement and the persistence of task completion." *Journal of applied behavior analysis*, vol. 49 1, pp. 122–37, 2016.

[58] S. Xu, Y. Sun, M. Huang, Y. Huang, J. Han, X. Tang, and W. Ren, "Emotional state and feedback-related negativity induced by positive, negative, and combined reinforcement," *Frontiers in Psychology*, vol. 12, 2021.

[59] P. L. Gunter and M. Coutinho, "Negative reinforcement in classrooms: What we're beginning to learn," *Teacher Education and Special Education*, vol. 20, pp. 249 – 264, 1997.

[60] F. D-DiGennaro, B. Martens, and L. L. McIntyre, "Increasing treatment integrity through negative reinforcement: Effects on teacher and student behavior," *School Psychology Review*, vol. 34, pp. 220 – 231, 2005.

[61] H.-H. Ma, "Comparison of the relative effectiveness of different kinds of reinforcers: A pem approach." *The behavior analyst today*, vol. 10, pp. 398–427, 2010.

[62] C. Chen, K. Z. K. Zhang, X. Gong, and M. K. O. Lee, "Dual mechanisms of reinforcement reward and habit in driving smartphone addiction," *Internet Res.*, vol. 29, pp. 1551–1570, 2019.

[63] Y. T. Wang, "Research on the user experience of infinite up-scrolling on chinese short video software douyin," *Lecture Notes in Education Psychology and Public Media*, 2023.

[64] D. Rozgonjuk, C. Sindermann, J. Elhai, and C. Montag, "Fear of missing out (fomo) and social media's impact on daily-life and productivity at work: Do whatsapp, facebook, instagram, and snapchat use disorders mediate that association?" *Addictive behaviors*, vol. 110, p. 106487, 2020.

[65] R. Muench and C. Muench, "Me without my smartphone? never! predictors of willingness for smartphone separation and nomophobia," *HCI International 2020 - Posters*, vol. 1226, pp. 217 – 223, 2020.

[66] F. Kaviani, B. Robards, K. Young, and S. Koppel, "Nomophobia: Is the fear of being without a smartphone associated with problematic use?" *International Journal of Environmental Research and Public Health*, vol. 17, 2020.

[67] A. Butt and T. Arshad, "The relationship between basic psychological needs and phubbing: Fear of missing out as the mediator." *PsyCh journal*, 2021.

[68] C. Chen, K. Z. K. Zhang, and S. J. Zhao, "Examining the effects of perceived enjoyment and habit on smartphone addiction: The role of user type," pp. 224–235, 2015.

[69] M. Zhitomirsky-Geffet and M. Blau, "Cross-generational analysis of predictive factors of addictive behavior in smartphone usage," *Comput. Hum. Behav.*, vol. 64, pp. 682–693, 2016.

[70] J. Wang and X. Zhang, "The reinforcements and punishments in social media addiction," *Journal of Education, Humanities and Social Sciences*, 2023.

[71] A. Darcin, S. Kose, C. O. Noyan, S. Nurmedov, O. Yilmaz, and N. Dilbaz, "Smartphone addiction and its relationship with social anxiety and loneliness," *Behaviour & Information Technology*, vol. 35, pp. 520 – 525, 2016.

[72] A. V. Deursen, C. L. Bolle, S. Hegner, and P. Kommers, "Modeling habitual and addictive smartphone behavior: The role of smartphone usage types, emotional intelligence, social stress, self-regulation, age, and gender," *Comput. Hum. Behav.*, vol. 45, pp. 411–420, 2015.

[73] L. Li, L. Wang, and X. Wang, "Effect of smartphone use before bedtime on smartphone addiction behaviors among chinese college students," *Frontiers in Psychology*, vol. 13, 2022.

[74] M. Anshari, Y. Alas, G. Hardaker, J. H. Jaidin, M. Smith, and A. D. Ahad, "Smartphone habit and behavior in brunei: Personalization, gender, and generation gap," *Comput. Hum. Behav.*, vol. 64, pp. 719–727, 2016.

[75] J. W. Adie, J. L. Duda, and N. Ntoumanis, "Autonomy support, basic need satisfaction and the optimal functioning of adult male and female sport participants: A test of basic needs theory," *Motivation and emotion*, vol. 32, pp. 189–199, 2008.

[76] M. Hyzy, R. R. Bond, M. Mulvenna, L. Bai, A. J. Dix, S. Leigh, and S. Hunt, "System usability scale benchmarking for digital health apps: Meta-analysis," *JMIR mHealth and uHealth*, vol. 10, 2022.

[77] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Q.*, vol. 13, pp. 319–340, 1989.

[78] E. Schmider, M. Ziegler, E. Danay, L. Beyer, and M. Bühner, "Is it really robust?" *Methodology*, 2010.

[79] M. J. Blanca-Mena, R. Alarcón-Postigo, J. Arnau, R. Bono Cabré, R. Bendayan *et al.*, "Effect of variance ratio on anova robustness: Might 1.5 be the limit?" 2017.

AI in the Detection and Prevention of Distributed Denial of Service (DDoS) Attacks

Sina Ahmadi
National Coalition of Independent Scholars, NCIS, Seattle, USA

***Abstract*—Distributed Denial of Service (DDoS) attacks are malicious attacks that aim to disrupt the normal flow of traffic to the targeted server or network by manipulating the server's infrastructure with overflowing internet traffic. This study aims to investigate several artificial intelligence (AI) models and utilise them in the DDoS detection system. The paper examines how AI is being used to detect DDoS attacks in real-time to find the most accurate methods to improve network security. The machine learning models identified and discussed in this research include random forest, decision tree (DT), convolutional neural network (CNN), NGBoosT classifier, and stochastic gradient descent (SGD). The research findings demonstrate the effectiveness of these models in detecting DDoS attacks. The study highlights the potential for future enhancement of these technologies to enhance the security and privacy of data servers and networks in real-time. Using the qualitative research method and comparing several AI models, research results reveal that the random forest model offers the best detection accuracy (99.9974%). This finding holds significant implications for the enhancement of future DDoS detection systems.**

Keywords*—*Artificial intelligence; Distributed Denial of Service (Ddos); machine learning; detection; accuracy

I. Introduction

In today's fast-paced digital landscape, web-based services and software have seen a significant rise, with approximately 57% of the global population now using the Internet [1]. While artificial intelligence (AI) and machine learning have become powerful tools across various industries, they have also introduced a host of security challenges, particularly in maintaining the performance and security of networks. Traditional networks often struggle to keep up with the demands for efficiency and robust security, leaving businesses vulnerable to cyber threats like Distributed Denial of Service (DDoS) attacks [2]. The escalating scale and frequency of these attacks highlight a critical problem: existing network infrastructures, including those managed by software-defined networking (SDN), are increasingly incapable of providing the security required to ensure smooth business operations. SDN, which manages network traffic through software platforms rather than hardware, provides a centralised control system that enhances network flexibility and manageability [3]. However, this centralised architecture also introduces vulnerabilities, particularly at the control layer, where attackers can disrupt or manipulate network traffic through DDoS assaults. These attacks are difficult to detect, and as their intensity and frequency continue to rise, they pose significant challenges for administrators and service providers in terms of identification and mitigation [4]. Although various machine learning methods have been proposed for detecting DDoS attacks, there remains a gap in understanding which of these methods is most effective in real-time scenarios, especially within SDN environments.

This study aims to fill this gap by evaluating and comparing different AI-based detection methods to determine the most accurate technique for real-time DDoS attack mitigation in SDN. By addressing the limitations of existing detection approaches, this research seeks to provide a more effective solution for identifying and preventing DDoS attacks. The findings will offer valuable insights into how AI can enhance the security of SDN infrastructures, contributing to the broader goal of protecting businesses from network disruptions and security breaches.

Ultimately, this study's significance lies in its potential to advance current cybersecurity measures by integrating AI into SDN environments. As the intensity and sophistication of cyberattacks continue to rise, finding more accurate detection methods becomes crucial. The results of this research will be essential for administrators and service providers seeking to safeguard their networks against DDoS attacks, thus ensuring greater operational stability and data security.

II. Literature Review

This section of the paper provides a detailed understanding of existing research and guides how this research presents a different perspective in the field. Numerous researchers have investigated AI-based detection methods to understand which method is most accurate for managing detective services.

A. Performance of AI / Machine Learning in DDoS Attack Detention

Meti et al., in their experiment, observe TCP traffic from actual networks along with the number of connected devices per second as an indicator [5]. In terms of precision, accuracy, and recall levels, the results of the comparison highlight that K-Nearest Neighbour (KNN) shows the best precision and accuracy. Zekri et al. suggested that a DT is also effective in the cloud network to detect DDoS attacks [6]. Sahoo et al. present an enhanced support vector machine (SVM) model that implements genetic algorithms (GA) and kernel principal component analysis (KPCA) [7]. Bakker et al. discuss the additional costs of using AI for DDoS attack detection in SDN [8]. Another study by Polat and co-authors confirms that KNN is the most accurate method for DDoS detection and security improvements [9].

B. Use of Machine Learning for DDoS Attack Detection

Huyu et al. present techniques for optimising and creating detection models by sending real-time traffic data to an offline learning network [10]. Data is collected through routers and transmitted to the offline pipeline for data transformation and feature engineering. Optimised models combined with existing models can be used to protect networks from DDoS attacks. Chayomchai et al.'s study focuses on the impact of cybercrime and DDoS attacks on banking institutions and how these institutions are responding to these negative effects [11]. The study claims that massive malware assaults target Indian banks, resulting in the theft of private customer information and huge financial losses. It also states that poorly designed detection models that do not need annotated data to supervise DDoS attacks can be improved using latent Dirichlet allocation (LDA). Another alternative is to use an extra classification layer to remove non-attack tweets from the dataset [12]. Ashraf and Latif propose a SOM-based solution to significantly improve accuracy [13]. However, SOM principles violate SDM principles as they are built on intelligence in the data plane [14]. Peng et al. provide a detection method for anomalous SDN streams in an SDN architecture [15]. They applied the same technique to detect DDoS attacks using DPTCM-KNN as the core algorithm. The results of this study demonstrate that the deployed technique is effective; however, the detection accuracy obtained needs improvement.

C. Machine Learning Models

1) Random forest: The random forest model of machine learning refers to a method of ensemble learning that combines forecasts from different DTs. This model is used for both regression and classification. Predictions from different decision trees are combined to formulate a final forecast, and every entry in a random forest contains a different subset of the data. By deploying ensemble techniques, the individual accuracy of DTs can be improved, making them more dependable for attack detection and prevention.

2) Decision tree: A DT refers to the graphical representation of a decision-making process that separates data based on the input values into different subgroups. Each subgroup produces further branching nodes that lead to other subgroups or outcomes. In regression and classification tasks, a DT is used to generate and present predictions based on data feature values. Decision trees are simple and convenient to understand, which is why they are helpful in detecting DDoS attacks. However, the accuracy of DTs needs improvements in terms of consistency in detection.

3) *Convolutional neural network:* A CNN is used for image classification and identification. Convolutional neural networks have innumerable applications, such as face recognition, image processing, object detection, and computer vision. They are AI-powered systems that use images as input to perform. Convolutional neural networks work automatically to learn certain features that might be used for categorisation. Algorithms are adjustable to different networks in SDN, making them suitable for environments where potential attacks might occur. In terms of detecting DDoS attacks, CNNs require large amounts of labelled data that might not be available in every scenario.

4) *NGBoosT classifier*: This machine learning model is used for tasks involving data classification. NGBoosT collects different predictions from trees and evaluates those predictions to propose a final prediction. It is helpful in detecting DDoS attacks because it provides predictions about their uncertainty, which is helpful in managing unclear and uncertain situations.

5) Stochastic gradient descent: It is a straightforward and efficient method to detect DDoS attacks on networks. Stochastic Gradient Descent is used to manage complex machine learning issues commonly occurring during text categorisation and language processing tasks. This method can also be applied to different linear models and is convenient for managing different DDoS detection scenarios.

D. DDoS

Xu et al. present a technique to detect DDoS attacks in SDN. The technique mainly depends on K-FKNN and K-means++. The proposed detection system would be implemented into the controller. The experimental results of this study reveal that the implemented technique is stable and efficient [16]. However, some drawbacks that make the technique less accurate include the longer time required to detect the attack and the high load it puts on the SDN resources.

Polat et al. present an alternative method for DDoS attack detection in SDN [9]. There are two aspects to the proposed detection system. The first aspect analyses the DDoS attack traffic and normal traffic on the SDN environment dataset. Filtration wrapping and feature selection methods were deployed in the second aspect to get the most effective features for machine learning model classification. However, there is a limitation to the introduced technique: the need for further enhancement in its performance and detection accuracy.

Novaes et al. have implemented a DDoS attack and a mitigation method in SDN [17]. The whole system is called LSTM-FUZZY. The detection system comprises three stages: characterisation, detection, and mitigation. The proposed system is ineffective due to its restricted scope for addressing the vulnerabilities of other networks. In addition, the model lacks the characteristics required to test different network topologies [17].

Sarwan et al. present a space- and time-efficient DDoS attack detection technique that possesses the characteristics of identifying hosts along with the attack's origin [18]. The technique uses different traffic characteristics to identify abnormal traffic behaviour. It also uses a threshold to identify normal and compromised hosts. This DDoS detection technique is efficient as it saves time and space. However, it does have the limitation of violating SDN standards by implementing logic into switches. Thus, there is a need to improve its performance and algorithmic accuracy.

Studies reveal that DDoS attacks are often released from a single host to seize or disable access by overloading the target network or system. The degree of damage or loss depends on the strength of the attacker's resources. There are different intentions and purposes for launching a Denial-of-Service

attack, which could be personal or institutional. Attackers use botnets or zombie computers to launch a DDoS attack. These attacks are pre-planned to disrupt or destroy a target network by using land moves and targeting a specific system [19].

Existing literature covers different dimensions of the proposed research topic; however, literature about the use of AI to detect and prevent denial-of-service attacks is minimal. Artificial intelligence is evolving in every industry to automate systems and enhance network performance. Researchers must disseminate knowledge on the use of machine learning applications in different fields to protect systems from malware and security assaults. The following section of the paper defines the research problem and the significance of understanding and addressing this problem.

III. Problem Definition

Denial-of-Service (DoS) attacks happen when authorised users fail to access network data due to malicious cyber threat activities launched by third parties. These attacks, driven by various motives, are launched against emails, passwords, databases, and websites to hack the private information of individuals or organisations. Business organisations using advanced software and computing networks to perform everyday operations are prone to such attacks. For example, banks store important organisational and customer information on online databases and transform sensitive information through networks that hackers can conveniently hack and misuse. DoS attacks create enormous challenges for administrators and managers in keeping operations streamlined. Organisations with poor detection infrastructure cannot detect abnormal traffic timely, resulting in attacks and network complications. Given the technological advancements and emerging threats, the major problem lies in the timely detection of DoS attacks and the application of accurate methods to network infrastructures.

Artificial intelligence has become an essential tool for transforming business experiences. It is revolutionising every industry by redefining traditional business practices and transforming customer experiences. By using AI applications and enhancing AI-based models, defence against DoS attacks is possible. For example, AI helps reduce the surface area vulnerable to attack, thereby minimising the options available to attackers. Load balancers mitigate this issue by restricting direct Internet traffic to specific parts of the network to avoid direct attacks. Similarly, Access Control Lists (ACLs) are useful in controlling which traffic would reach applications in a given time [20].

A. DDoS Detection and Prevention

Common mechanisms for detecting and preventing DDoS include attack detection, prevention, and reaction. However, it is difficult to detect DDoS attacks in a network as it is hard to differentiate between abnormal and normal network traffic during ongoing operations. The detection of abnormal traffic in a network is the first step to detecting DDoS attacks. In addition, AI classification methods can help in identifying good and bad packets. Bed packets labelled as abnormal traffic would be dropped. The number of packets, time interval variance, average size of packets, number of bytes, size variance, and packet rate are common characteristics that help differentiate between good and bad packets.

B. Artificial Intelligence Techniques

Most relevant AI techniques include machine learning, natural language processing, and speech recognition. Machine learning algorithms are utilised in most of the settings. These techniques, including Naive Bayes, support vector machines, and neural networks, are implemented based on the nature and frequency of attacks [21].

C. Trends of DDoS Attack

In the fast-paced business world, DDoS attacks are commonplace. For instance, public networks experience frequent instances of high-intensity floods, which significantly affect the normal flow of network traffic and disrupt normal functioning. Although the security protocols of a network protect it against DDoS attacks, trends of attack vary based on the strength of the security protocols of a network. The number of organisations experiencing DDoS attacks is increasing annually, along with the growing dependence on software and databases for managing organisational processes. The integration of technology is simultaneously easing and complicating institutional processes.

D. Integration of AI Models into Networks to Prevent DDoS Attacks

Distributed Denial of Service attacks have been demonstrated to be major threats to the Internet, causing major losses to organisations and governments. With the advancement of technology, it has become convenient for attackers to launch DDoS attacks at low costs. Attackers use unknown hosts or computers to launch DDoS attacks, and it is hard to detect them without having advanced security infrastructures embedded in AI models. However, different AI models, like machine learning algorithms, are available to help detect DDoS attacks. These models vary in accuracy and performance and can be used based on network settings and requirements to prevent DDoS attacks. Fig. 1 shows a machine learning-enabled DDoS detection architecture.

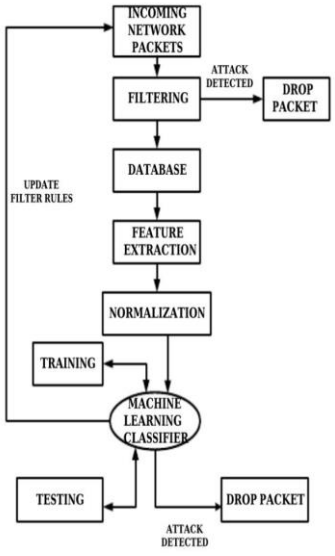


Fig. 1. DDoS attack detection architecture based on machine learning.

E. Accuracy and Adaptability

It is essential for AI-based detection models to provide accurate detection results to protect networks against unauthorised access and use. Models with the greatest accuracy rates are adaptable under certain standards. Detection models should be able to predict possible abnormalities in the network and must inform administrators to timely mitigate those abnormalities. AI-powered detection models can prevent DDoS attacks and protect networks by focusing on security and privacy [22]. The accuracy of machine learning detection models can be determined using Eq. (1).

$$Accuracy = Number\ of\ Correct\ Predictions / Total\ Number\ of\ Predictions \quad (1)$$

IV. Methodology / Approach

A. Research Design

This study uses qualitative research methodology to investigate several AI models and employ them in the DDoS detection system. Using qualitative research methodology, this study aims to explore the application of AI to detect DDoS attacks in real time in banks to find the most accurate methods to improve network security. Qualitative research methodologies encompass the use of theory and literature to explore the diverse perspectives, experiences, and behaviours of people. This approach is carefully chosen to explore and compare machine learning detection models to detect and prevent DDoS attacks. Conducting a qualitative analysis of existing literature helps in understanding the different available methods and their accuracy. This analysis may assist in determining which method is most accurate to avoid network disturbances and breaches. The objective of this study is to compare detection models based on their performance and accuracy scores and propose further improvements to enhance the security of organisational and institutional networks.

B. Research Setting and Participants

The research setting consists of private banking institutions. The participants are banks that have experienced DDoS attacks in the previous two years and used different detection and prevention techniques to prevent further attacks. They also include all researchers who presented relevant experiences regarding AI-detection methods and accuracy measures. Banks are chosen in order to study the use of AI in the detection of DDoS in real-time. Moreover, AI-powered methods are effective at detecting these attacks and maintaining security standards.

C. Data Collection

The data collection method started with the use of Google Scholar, wherein different resources were collected using several research terms such as 'Denial of Service Attacks,' 'Applications of AI in detecting network attacks,' 'network security,' and 'ML models of detection.' These search terms were chosen to retrieve relevant results from the search engine. Different articles were reviewed to collect relevant information and then compared with the observations gained from the AI detection models implemented in banks to avoid DDoS attacks. The comparison aims to understand how different models work in theory and practice as well as how they could be improved further to detect and prevent DDoS attacks in various institutional settings. Through this comparative analysis strategy, the research aimed to collect and evaluate diverse perspectives and experiences to inform its findings and conclusions. Research results would contribute to the existing literature by presenting a novel dimension of DDoS and network attacks.

D. Data Analysis

The thematic analysis approach was utilised to analyse data collected from the literature review. Resources were carefully selected and analysed to guide research problems. DDoS attacks are a wide research topic that researchers have extensively explored from different dimensions. It was ensured that the proposed research problem addressed a novel concern, and relevant literature was used to guide the research. The analysis process was detailed and comprehensive to enhance the study's credibility and guide research conclusions. Data collected from the literature review and banks using AI models to detect and prevent DDoS attacks was analysed collectively to guide discussions and the research conclusion. Three banks were randomly selected to provide e-banking services. Real-time e-banking transactions were checked to track fraudulent activities. An in-depth analysis of e-banking transaction logs was conducted. Banks employed a multi-layered security approach to prevent DDoS attacks, including the use of one-time passwords to ensure that authorised users have access to their accounts. Data from banks using AI-powered solutions to ensure their data is protected and the minimal probability of attacks was analysed. Two banks use blockchain-powered DDoS mitigation strategies and solutions to prevent DDoS attacks. Analysis of real-time e-banking transactions of banks revealed that AI and machine learning models/applications provide real-time protection to online transactions and ensure that institutional and customer data is protected from unauthorised access or breaches. The analysis of the collected data aimed to provide an in-depth understanding of Internet networks prone to DDoS attacks, the nature of attacks, causes and effects, methods to detect and prevent them, and ways in which these methods could be further improved. Furthermore, the correlation between research objectives and themes is also discussed. It highlights how information gained from these themes guides understanding of the research problem and convinces the need for improved security protocols to prevent future DDoS attacks. Thematic analysis not only summarises research findings but also contextualises these findings for readers and future researchers.

E. Ethical Considerations

Research ethics were followed throughout the research process. All resources utilised during the literature review were properly acknowledged through appropriate citations and references, giving credit to the authors. The literature analysis was presented without any personal amendments or changes. All three banks approached to investigate how AI is helpful in detecting DDoS attacks in real-time requested not to reveal their identities in the research paper. In order to respect their privacy, their identities have been concealed, and all discussions pertaining to them have been conducted anonymously. The study addresses biases and conflicts observed in research studies. It ensures that personal conflicts

and biases are avoided throughout the paper to ensure the generalisability of the results and maintain integrity throughout the research process.

V. RESULTS

A. Machine Learning Models

This research paper examined and compared five machine learning models: random forest, DT, CNN, NGBoosT classifier, and SGD. Different performance parameters were used to compare the precision and accuracy of each machine learning model. The accuracy score of each model is illustrated in Fig. 2.

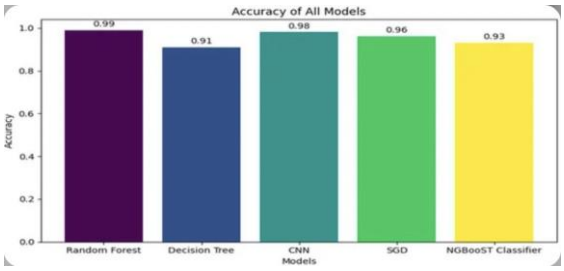


Fig. 2. Accuracy of all models.

Accuracy is a quantitative measure that quantifies the ratio of false negatives and positives to the terms present in the numerator. The numerator further specifies the sum of true negatives and true positives. Eq. (2) defines accuracy as follows:

$$Accuracy = ((TP + NP) / (FP + FN)) * 100 \quad (2)$$

The accuracy scores were used to measure the effectiveness of the five machine learning models in detecting DDoS attacks. Fig. 2 presents the ideal accuracy score of 0.99 for the random forest machine learning model. The purple bar of the random forest model shows how accurate the prediction would be compared to the rest of the four machine learning models. Similarly, CNN, SGD, NGBoosT classifier, and DT show accuracy scores of 0.98, 0.96, 0.93, and 0.91, respectively. Hence, based on the accuracy scores, the random forest machine learning model is the most effective for implementation in institutional settings, whether banks or other organisations, to detect and prevent DDoS attacks [23].

B. Comparative Analysis of Machine Learning Models

A comparative analysis of five machine learning models reveals that each model has its strengths and weaknesses. The random forest model is effective in making predictions for complex settings and is conveniently adjustable in different SDN settings compared to DT and NGBoosT classifiers. The random forest model offers flexibility in different domains compared to NGBoosT, which cancels overfitting into multiple settings. All five machine learning models differ in their approach to handling DDoS attacks in different settings. The accuracy and precision rate of each model differ, highlighting the usefulness and performance of each model in institutional settings. Each model can be individually deployed in different institutional settings based on the security requirements and infrastructure of that institution [24].

VI. DISCUSSION

A. DDoS Attack Detection and Prevention in Banking Industry

An in-depth analysis of the e-banking transaction logs of three banks reveals that AI applications are helpful in detecting and preventing DDoS attacks in real time. In banking institutions, blockchain-powered DDoS mitigation strategies are utilised to enhance security standards and protocols. Banking institutions are implementing strong security protocols and exploring new defences against DDoS attacks. They are using AI models to build strategies for identifying the origins and underlying causes of attacks. Eliminating traditional DDoS prevention approaches, banks are deploying advanced security measures to avoid financial and non-financial losses. Machine learning and AI have made it possible to automatically detect and prevent DDoS attacks [25].

Based on the analysis of existing detection and prevention models implemented by banks, some changes are proposed. With the following structure, the random forest model would provide the most accurate predictions about DDoS attacks.

The proposed changes in the DDoS detection and prevention models would yield outstanding outcomes and protect real-time customer interactions and transactions. This is depicted in Fig. 3.

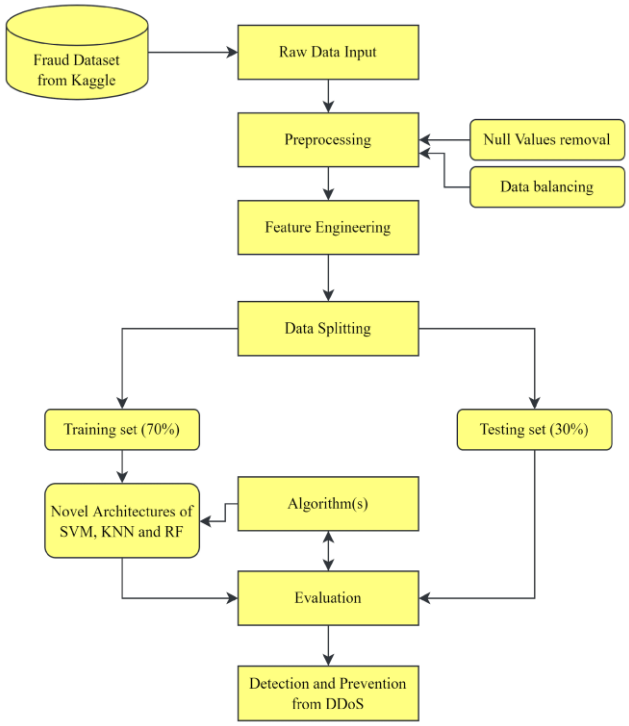


Fig. 3. Proposed machine learning model using e-banking datasets.

B. Adopting Advanced AI Solutions to Detect and Prevent DDoS Attack

In the context of cyberattacks and threats, organisations should understand the importance of advanced and updated AI models to detect and prevent DDoS attacks. Along with advancements in machine learning and AI, cyber threats are

evolving. Therefore, it is essential to timely understand the need for improved machine learning models and integrate them as needed to avoid network breaches and data thefts. Organisations and administrators can focus on evolving AI trends and models being deployed to prevent DDoS attacks. Based on the accuracy code and analysis of the machine learning models used in this study, the random forest model is most suitable for predicting DDoS attacks.

C. Privacy and Confidentiality

Banks are dealing with highly sensitive customer information that could cause major complications if breached. Banks have a paramount responsibility to protect their customers' privacy and confidentiality. This can only be achieved by using upgraded security protocols. Literature analysis reveals that corporations and organisations use different detection and prevention mechanisms to avoid financial and non-financial losses due to DDoS attacks. In dealing with customers' private information, organisations such as banks are expected to adopt ever-changing AI applications that protect network traffic and ensure no abnormalities exist. Collaborative DDoS attack detection and prevention methods can be used to detect attacks at different locations before they occur.

D. Implementation Challenges

The implementation of a machine learning model to protect organisational networks against DDoS is a challenging task. Organisations can protect their networks against these attacks by deploying advanced machine learning and AI-powered solutions and technologies. One of the biggest challenges is that machine learning and AI require a large amount of labelled data to make accurate predictions. Obtaining detailed data in any SDN setting is challenging, which limits the adaptability of these models to effectively detect and prevent DDoS attacks in every situation. Another challenge is resource allocation. The deployment of machine learning models requires heavy investments, imposing financial burdens on organisations. Moreover, not all organisations can afford the cost. In addition to purchasing and implementing machine learning models, regular updates and space requirements are also a concern. Organisations spend to mitigate cyberattacks; however, it is expensive to deploy machine learning models as they require planning from resources to training for successful deployment. To cut cyberattack costs permanently, organisations should consider and prioritise the implementation of machine learning models according to their needs and security infrastructure. This will ensure they have strong security protocols and that their assets are protected from unauthorised access and theft.

Financial institutions, such as banks, are at great risk because of the monetary value of their data. Based on our analysis, the random forest machine learning model is one of the most effective models for making accurate predictions. By enhancing its adjustability, this model can be implemented in financial organisations to timely detect and prevent DDoS attacks. The random forest model can detect abnormal traffic in a network to make predictions regarding the occurrence of a DDoS attack at a certain point. Financial institutions can effectively implement this model to avoid additional expenses associated with addressing cyberattacks and data theft [26].

VII. Conclusion

The Internet has transformed the world over the years. It has become an essential need for every institution and organisation. In this fast-paced world, the Internet is transforming lives and complicating things as well. Cyberattacks have become commonplace in the technically advanced world, wherein an unknown person can assess the private information of institutions or organisations. A DDoS attack is a form of cyberattack in which hackers take control of the central controller to disrupt or completely damage a network. Hackers can launch this attack from any server. The detection and prevention of DDoS attacks is a complicated and sometimes impossible task.

This paper aims to investigate the role of AI in the detection and prevention of DDoS attacks in real-time. To achieve this objective, data was collected and analysed from existing literature and three anonymous banks. Data analysis reveals that institutions are using attack mitigation techniques; however, these techniques are not as effective as machine learning models in defending against DDoS attacks. From the reviewed machine learning models, the random forest model is the most effective model to implement in organisational settings to detect and prevent DDoS attacks promptly. Considering the frequency of cyberattacks, researchers strongly advocate for the integration of machine learning models into the strategic planning of large organisations and institutions. This will help mitigate long-term challenges by ensuring they implement advanced security protocols. It is recommended that, with evolving technologies, organisations understand the dynamic need for security upgrades to compete efficiently in the industry.

VIII. Future Work

This research discussed five machine learning models and found that the random forest model makes more accurate predictions than the other models. Future researchers can conduct further work on the accuracy and effectiveness of the random forest model and why it is most accurate compared to other models. Future research can highlight how machine learning models are making a difference compared to traditional models in detecting and preventing DDoS attacks. Researchers can also explore how the random forest model effectively detects DDoS attacks in different network settings. The adjustability, accuracy, and precision of the random forest model can be studied in detail to justify its effectiveness in comparison to other models.

Moreover, advancements in prevention and detection techniques and mechanisms can transform organisational outcomes. Hence, by integrating and adopting advanced machine learning models to detect DDoS attacks, organisations can safeguard customers' privacy and confidentiality and streamline their everyday operations.

References

[1] Internet growth usage statistics [Internet]. 2019 [cited 2024 Jan 10]; Available from: https://www.clickz.com/internetgrowthusage-stats-2019-time-online-devices-users/235102/

[2] Singh J, Behal S. Detection and mitigation of DDoS attacks in SDN: A comprehensive review, research challenges and future directions. Comput. Sci Rev 2020;37:100279.

[3] Sheikh MNA, Hwang IS, Ganesan E, et al. Performance assessment for different SDN-based controllers. In: Proceedings of the 2021 30th Wireless and Optical Communications Conference (WOCC), Taipei, Taiwan; 2021: p. 24-5.

[4] Wang Y, Wang X, Ariffin MM, et al. Attack detection analysis in software-defined networks using various machine learning methods. Comput Electr Eng 2023;108:108655.

[5] Meti N, Narayan DG, Baligar VP. Detection of distributed denial of service attacks using machine learning algorithms in software defined networks. In: Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Manipal, India; 2017: p. 1366-71.

[6] Zekri M, El Kafhali S, Aboutabit N, et al. DDoS attack detection using machine learning techniques in cloud computing environments. In: Proceedings of the 2017 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech), Rabat, Morocco; 2017: p. 1-7.

[7] Sahoo KS, Tripathy BK, Naik K, et al. An evolutionary SVM model for DDOS attack detection in software defined networks. IEEE Access 2020;8:132502-13.

[8] Bakker JN, Ng B, Seah WK. Can machine learning techniques be effectively used in real networks against DDoS attacks? In: Proceedings of the 2018 27th International Conference on Computer Communication and Networks (ICCCN), Hangzhou, China; 2018: p. 1-6.

[9] Polat H, Polat O, Cetin A. Detecting DDoS attacks in software-defined networks through feature selection methods and machine learning. Sustain 2020;12:1035.

[10] Huyn J. A scalable real-time framework for DDoS traffic monitoring and characterization. In: Proceedings of the Fourth IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, Austin, TX, USA: 2017: p. 265-6.

[11] Mhamane SS, Lobo LMRJ. Internet banking fraud detection using HMM. In: Proceedings of the 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12), Coimbatore, India; 2012.

[12] Chayomchai A, Phonsiri W, Junjit A, et al. Factors affecting acceptance and use of online technology in Thai people during COVID-19 quarantine time. Manag Sci Lett 2020;10:3009-16.

[13] Ashraf J, Latif S. Handling intrusion and DDoS attacks in software-defined networks using machine learning techniques [Internet]. IEEE Xplore. 2014. p. 55-60. Available from: https://ieeexplore.ieee.org/abstract/document/6998241

[14] 2014 IEEE National Software Engineering Conference [Internet]. Interdisciplinary Centre for Mathematical and Computational Modelling. 2014 [cited 2024 Aug 6]. p. 55-60. Available from: https://www.infona.pl/resource/bwmeta1.element.ieee-conf-000006979384/

[15] Peng H, Sun Z, Zhao X, et al. A detection method for anomaly flow in software defined network. IEEE Access 2018;6:27809-17.

[16] Xu Y, Sun H, Xiang F, et al. Efficient DDoS detection based on K-FKNN in software defined networks. IEEE Access 2019;7:160536-45.

[17] Novaes MP, Carvalho LF, Lloret Jaime, et al. Long short-term memory and fuzzy logic for anomaly detection and mitigation in software-defined network environment. IEEE Access 2020;8:83765-81.

[18] Ali S, Alvi MK, Faizullah S, et al. Detecting ddos attack on SDN due to vulnerabilities in openflow. IEEE 2019 International Conference on Advances in the Emerging Computing Technologies (AECT), Al Madinah Al Munawwarah, Saudi Arabia; 2020: p. 1-6.

[19] Raza MS, Sheikh MNA, Hwang I, et al. Feature-selection-based DDOS attack detection using AI algorithms. Telecom 2024;5:333-46.

[20] AWS. What is a DDOS attack & how to protect your site against one [Internet]. Amazon Web Services, Inc. 2024 [cited 2024 Aug 6]; Available from https://aws.amazon.com/shield/ddos-attack-protection/

[21] D G. DDoS detection and prevention based on artificial intelligence techniques. Sci Bull Nav Acad 2019;22:134-43.

[22] Bortey L. How do you measure machine learning model accuracy after data preprocessing? [Internet]. Linkedin. 2023 [cited 2024 Aug 6]; Available from: https://www.linkedin.com/posts/loretta-bortey-b2517481_how-do-you-measure-machine-learning-model-activity-7095092695324311552-KtJM

[23] Islam U, Muhammad A, Mansoor R, et al. Detection of distributed denial of service (DDOS) attacks in IOT based monitoring system of banking sector using machine learning models. Sustain 2022;14:8374.

[24] Noi PT, Kappas M. Comparison of random Forest, K-Nearest Neighbor, and Support vector machine classifiers for land cover classification using Sentinel-2 imagery. Sensors 2017;18:18.

[25] D G. DDoS detection and prevention based on artificial intelligence techniques. Sci Bull Nav Acad 2019;22:134-43.

[26] Zhang C, Liu C, Zhang X, et al. An up-to-date comparison of state-of-the-art classification algorithms. Expert Syst Appl 2017;82:128-50.

A Hybrid Regression-Based Network Model for Continuous Face Recognition and Authentication

Bhanu Kiran Devisetty*, Ayush Goyal, Avdesh Mishra, Mais W Nijim, David Hicks, George Toscano
Department of Electrical Engineering and Computer Science, Texas A&M University-Kingsville, Kingsville, Texas

***Abstract*—This research proposes a continuous remote biometric user authentication system implemented with a face recognition model pre-trained on face images. This work develops an algorithm combining the Hybrid Block Overlapping KT Polynomials (HBKT) and Regression-based Support Vector Machine (RSVM) methods for a face recognition-based remote user authentication system that uses a model pre-trained on the ORL, Face94 and GT datasets to recognize authorized users from face images captured through a webcam for continuous remote biometric user authentication. HBKT polynomials enhance feature extraction by capturing local and global facial patterns, while RSVM improves classification performance through efficient regression-based decision boundaries. The system can continuously capture user face images from the user's webcam for user authentication, but it can be affected by lighting variations, occlusion, and computational overhead from continuous image capture. This has been implemented in a Python program. The proposed method, when compared to previous state-of-the-art algorithms, was observed to have higher F-measure, accuracy, and speed, for most of the cases. The proposed method was observed to have accuracies of 98.82% (ORL dataset), 96.73% (GT dataset), and 95.9% (Face94 dataset).**

***Keywords*—Vision-based computing; object detection; face detection; face recognition; feature extraction; feature coefficients; classification; authentication; biometrics; biometric authentication**

I. INTRODUCTION

Before allowing a user access to a secured location, resource, or system, authentication confirms the user's identity using specific credentials. Generally speaking, authentication only occurs during the user's initial interaction with the system [1]. Under these circumstances, the user is prompted with knowledge-based authentication that resembles password inquiry. The author is deemed authenticated if the user correctly replies with an example password [2]. However, one-time authentication is the standard for authentication; researchers have examined a variety of concerns, including security flaws and user dissatisfaction. For instance, when trying to unlock a smartphone with a pattern- or password-based authentication system, the user must concentrate on multiple authentication processes [3]. Driving while distracted is one concern that could arise from this for user safety. One-time authentication's lack of ability to identify attackers after the initial authentication is a significant security vulnerability [4]. For instance, if an authorized user forgets to log out or leaves their authenticated device unattended, an unauthorized user may be able to access their private resources. The above mentioned issues have prompted research into continuous authentication techniques [5]. To ensure the person who used the device to authenticate themselves the first time is still using it, these methods keep an eye on how the user interacts with it even after logging in. Simple security measures that lock users' devices and require them to re-enter their passwords after a certain amount of inactivity served as the foundation for the first attempts in this approach [6]. Though there is potential for improvement, such systems may upset people while they present a window of vulnerability. There are significant increases in the research literature on using behavior metrics, such as keystroke frequency, and biometrics, such as facial features, for continuous authentication [7]. This field has seen rapid growth, as evidenced by the numerous publications in the last ten years.

Researchers' interest in wearable medical sensors (WMSs) has grown, and these devices are beginning to be used in clinical settings. These sensors measure biological data, such as body temperature, blood pressure, and heart rate. The sales of 33 million wearable health monitoring gadgets, according to a recent Business Insider piece, are also analyzed [8]. According to the forecast, this number will rise quickly after reaching 148 million in 2019. The author proposes that such biomedical signals can also be utilized to provide authentication since they will be gathered in any case for health monitoring purposes [9]. Three factors make using biological data continuously collected for user identification and verification seem promising [10]. First, there is no need for an additional device on the body to use biological signals for authentication if WMSs are already gathering them for medical purposes. Secondly, the user is not heavily involved in collecting this data, as it is done transparently. In contrast to conventional biometrics and behavior metrics such as facial features and keyboard patterns, which may sometimes become unavailable, the biomedical signal stream obtained from wearable medical devices (WMSs) remains accessible as long as the user wears them [11] – [15]. The challenges encountered in the existing approaches are resolved using the anticipated model discussed below:

The transparent continuous authentication system H-RSVM is presented in this paper. It is based on a set of facial images provided to validate the user's authenticity. It is generally called the face recognition process. Image samples are continuously collected from the available online resources for diagnostic and therapeutic reasons. The primary distinction between a biometric characteristic and an input image is that the latter lacks the necessary discriminatory capacity to identify individuals independently. Therefore, it is doubtful that an authentication decision based solely on facial images will be adequately discriminative. Still, integrating many image samples into a single framework makes for a powerful continuous authentication mechanism. This study improves the biometric

remote authentication using HBKT polynomials integrated with RSVM, and makes it faster and more accurate than earlier studies. It affords a smooth, uninterrupted user authentication through the actual capture of images by webcams for real time and non-intrusive authentication. The proposed method is very stable and optimized for use in applications like secure remote authentication. The following sums up our primary accomplishments:

1) For the design of any continuous authentication system, this work proposes a novel H-RSVM where the features from the image datasets are analyzed and classified to have better prediction accuracy.

2) To assess the ability of prediction and to differentiate individuals, these works propose implementing a continuous authentication system known as H-RSVM and analyzing its accuracy and scalability.

3) We present a method for adaptive authorization and discuss its potential application in resolving challenges faced by users due to possible erroneous restrictions imposed by continuous authentication systems. This work outlines several potential defences against the suggested continuous authentication method and many ways to avoid them.

The work is structured as follows: Section II gives an extensive analysis of various prevailing approaches used for authentication during face recognition. Section IV delves into the problem statement. The methodology is elaborated in Section IV. The numerical outcomes are discussed in Section V, and work summarization is provided in Section VI.

II. Literature Review

It has been suggested in numerous research studies that biometrics be used for ongoing user authentication. In real-time authentication, inertial data demonstrates location, movement and device orientation concerning its surroundings [16]. Based on user activity, nonintrusive authentication techniques utilize this data to generate behavioral characteristics, such as voice, hand gestures, keyboard patterns, locomotion, touchscreen operations, and signature movements. Yu et al. [17] were the first to use a one-class distance-based classifier and build a sizable dataset for continuous authentication. They combined touchscreen, pressure, acceleration, time interval, and touch area size data with inertial data from the device's gyroscope and accelerometer. Their objective was to discern the authenticity of smartphone users, distinguishing between genuine and fraudulent individuals [18]. By generating research profiles, they achieved an Equal Error Rate (EER) of up to 3.6%, which varied depending on how each user held their device while entering their Personal Identification Number (PIN). By examining keyboard and handwriting data gathered from password submissions, Sun et al. [19] validated smartphone users in their study using deep learning techniques. The researchers categorized the images using various models, such as Naïve Bayes, Bayesian Net classifiers, and multilayer perceptrons (MLP) [20].

In a subsequent study, Menotti et al. [21] utilized a comprehensive dataset from Google's Abacus project, employing advanced techniques such as time-based deep feature extraction. Their research's main goal was to develop user authentication, and they did so by using recurrent and convolutional neural networks (CNNs and RNNs, respectively). The Google Abacus dataset, which includes 1500 people's information in their natural settings, was used for this. It is crucial to remember that this dataset is not publicly accessible. The researchers used the D-RNN model to categorize the data. Qin et al. [22] used the HMOG dataset, which included more than 27,000 samples from 10 participants, to show notable progress in the field. Wavelet, frequency, and time domain features were used to test algorithms like SVM, K-Nearest Neighbors and Hidden Markov Models. Based on an adaptation of Zhu et al. [23], speech, keyboards, touchscreens, gestures, handwriting, and mobility have all been studied in research. The requirement for user input during the whole authentication procedure is the primary disadvantage of gesture-based research [24]. It is impossible to identify an unauthorized user after the device is unlocked. However, compared to other methods, keystroke dynamics solutions are more constrained and require more data since they are impacted by fluctuations in user behavior, such as moods [25].

Moreover, changing keyboards may cause previously taught actions to become erratic. Additional research constraints conducted using touchscreens include the variability of interactions based on the device's orientation and the influence of user behavior on the level of interactivity [26]. Handwriting-based methods are seldom designed to support continuous authentication due to the smartphone's inability to detect pattern changes accurately. Speech-based authentication technologies are also hindered by background noise from the surrounding environment [27]. Furthermore, gait-based recognition, which relies on maintaining consistent positioning of the body's sensors, is vulnerable to modifications in walking patterns resulting from adjustments in clothing. Various projects have been dedicated to verifying human actions in a specific work context, such as retrieving a phone from a surface, initiating a phone call, or inputting a password [28]. While a more straightforward machine learning problem might yield better outcomes, it needs more consistency. For instance, the phone only knows how to classify various behaviors once the same activity is performed frequently once a user has been granted permission. Two different user activities are detectable by motion sensors. While the second is complex, the first is easy. Simple activities include sitting, sleeping, walking, climbing and descending stairs, and lying down [29]. In contrast, complex tasks involve riding a bike, exercising, changing clothes, and operating a vehicle. A method for determining when people are driving, biking, walking, using the bus, or utilizing the train in real time was created by Vijaya et al. [30]. This technique makes use of accelerometer and GPS data. They demonstrated that GPS data processed using principal component analysis (PCA) and recursive feature elimination (RFE) in conjunction with a random forest (RF) classifier yielded 96% accuracy. However, because GPS requires user clearance to access data and has high battery consumption, its usage in CA contexts is not practicable for real-world deployment. Support vector machine (SVM) models using accelerometer and gyroscope data were developed [30]. The model they developed was able to identify 95% of the different walking patterns, including going upwards (79%), standing (92%), sitting (94%), lying down (100%), and climbing

stairs (72%) [31] – [35]. Similar methods were used to investigate human activity detection with artificial neural networks and deep learning, yielding a 95% accuracy rate.

III. PROBLEM STATEMENT

The challenge in continuous user authentication is to create effective and non-intrusive method to distinguish between genuine and fraudulent users in various real-world conditions. Existing methods such as Gesture based, keystroke dynamics, and speech-based authentication continue to face challenges of variation, noise or changes in orientation thus producing varying results. Further, gait-based recognition is vulnerable to cloth variation or change of body posture whereas the GPS based systems though very accurate are not feasible due their high-power consumption and privacy issue. Thus, the problem is to develop a continuous authentication framework that combines multiple behavioral and biometric data and real-time authentication with minimal user input. The proposed method provides better optimal accuracy and flexibility by utilizing combined behavioral parameters and biometrics to enhance security to particular contexts of use without compromising the convenience of the end-users.

IV. METHODOLOGY

Face recognition is required in various applications, and significant progress has been identified in this research. The proposed work of remote user access with face verification consists of the following stages. Samples of face images from databases and the overall architecture are shown and discussed.

A. Dataset

The 400 grey-scale pictures in the AT&T database, once called "The ORL Database of Faces," feature 40 subjects [36-37]. Each issue has ten pictures that include every potential combination of attributes. Each subject's face samples are provided in the Portable Gray Map (PGM) format. Fig. 1 displays a variety of face samples from the databases that are utilized for training the model used in this work.

B. Pre-processing

It is the most widely utilized method for completing pre-processing. It changes the values in the image to fall between 0 and 1. Zero means and z-score normalization are the normalization techniques taken into consideration here, as stated in the equation below:

$$X_i' = \frac{X_i - mean\ (X)}{SD\ (X)} \tag{1}$$

In this case, the symbol X_i' represents data that has been normalized. The average value of the input X is represented by the mean (X), and the SD (X) represents the standard deviation. The equation below gives the expression for computing standard deviation:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - mean\ (X)\right)^2} \tag{2}$$

In this case, the provided input values' standard deviation is indicated as σ.

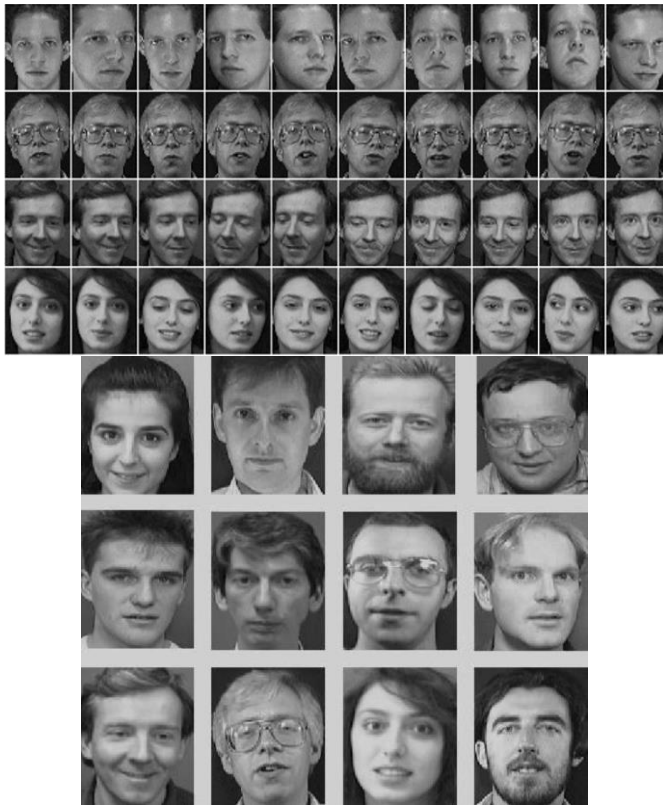


Fig. 1. Samples from the face image databases [36-37].

C. Feature Extraction

The weighted and normalized Krawtchouk polynomial's nth order is defined in the equation below as follows:

$$K_n\ (x; p, N-1) = \sqrt{\frac{\omega_K(x)}{\rho_K}}\, F_1\left(-n, -x;\ -N+1;\ \frac{1}{p}\right)$$

$$\text{where, } n, x = 0,1, \ldots, N-1;\ p \in (0,1) \tag{3}$$

Fig. 2 shows the overall architecture and functionality of the proposed work, which includes both feature extraction using HBKT polynomials and face recognition using RSVM.

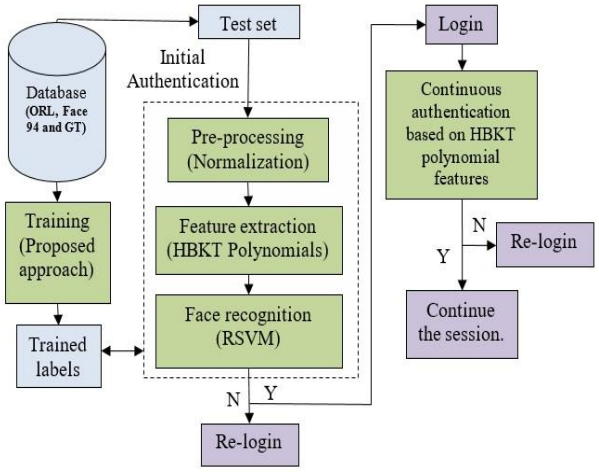


Fig. 2. Overall architecture of proposed work.

The norm and weight functions of the KP, denoted by ρ_K and ω_K, have been defined in the following manner:

$$\omega_K(x) = \binom{N-1}{x} p^x (1-p)^{N-x-1} \quad (4)$$

$$\rho_K(n) = (-1)^n \binom{1-p}{p}^n \frac{n!}{(-N+1)_n} \quad (5)$$

The three-term recurrence method is utilized to estimate KP coefficients since it is more computationally expensive to calculate KP coefficient values employing gamma and hypergeometric functions. Numerous studies have looked into this strategy. The author offered initial recurrence relation in n −direction and established recurrence relation. The recurrence relations divide the n- x plane into two regions. One of these regions has its KP coefficients computed, while the other region's coefficients are derived using a symmetry relation.

The author suggested using a dual n −direction recurrence relation involving forward and backward computations to find new KP coefficients. KP plane comprises four triangular sections for which coefficients are determined for two of the triangles. In contrast, the remaining two triangles' coefficients are derived using symmetry relation. To compute KP coefficients, a rapid recurrence relation was recently developed. As shown in Fig. 3, this entails partitioning KP space into sections designated by both primary and secondary diagonals.

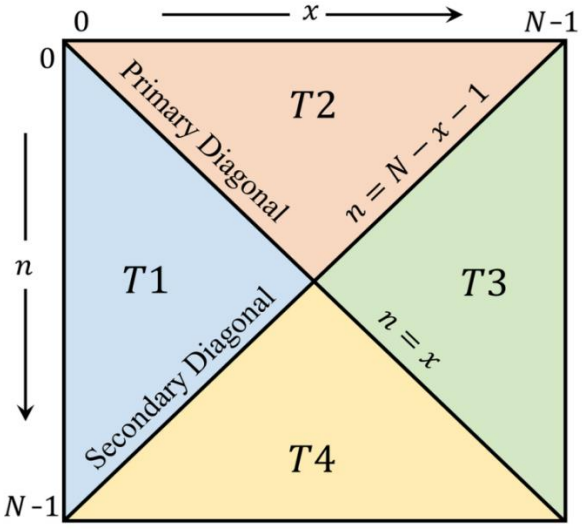


Fig. 3. KP plane.

For a wide range of parameters p, this method was better regarding computing speed, accuracy, and handling large signal quantities. In light of this rationale, the algorithm proposed is examined in this investigation. The procedure for computing the KP coefficients, denoted as $K_n(x;\ p, N-1)$ for brevity, is outlined in Fig. 4 (hereafter referred to as $K_n(x)$).

1) First, we compute $K_n(0)$ and $K_n(1)$ as follows:

$$K_n(0) = \sqrt{\frac{(N-n)p}{n(1-p)} * K_{n-1}(0)} \quad (6)$$

$$K_n(1) = \frac{-n+p(N-1)}{p(N-1)} \sqrt{\frac{(N-1)p}{(1-p)}} K_n(0) \quad (7)$$

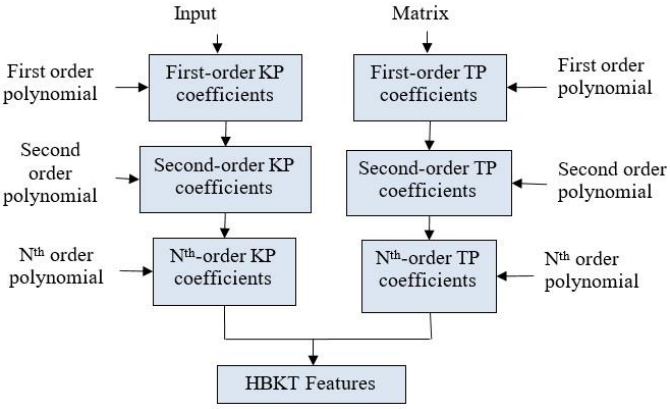


Fig. 4. HBKT feature extraction.

2) The recurrence relation in $the\ n$ direction can be used to determine KP coefficients within the temperature range T1.

$$\gamma_1 K_n(x+1) = \gamma_2 K_n(x) + \gamma_3 K_n(x-1) \quad (8)$$

$$\gamma_1 = \sqrt{p(N-x-1)(1-p)(x-1)} \quad (9)$$

$$\gamma_2 = -n + p(N-x-1) + x(1-p) \quad (10)$$

$$\gamma_3 = \sqrt{x(1-p)p(N-x)} \quad (11)$$

3) Using the symmetry connection about the main diagonal at n = x, the coefficients within the temperature range T2 can be found.

$$K_n(x) = K_x(n) \quad (12)$$

4) The symmetry relation concerning the secondary diagonal at $n = N - x - 1$ can be used to find the coefficients inside the temperature ranges T3 and T4.

$$K_{N-x-1}(x) = (-1)^{N-n-x-1} K_n(x) \quad (13)$$

5) The KP coefficients for p more significant than 0.5 values are computed using the equation below.

$$K_n(x: 1-p) = (-1)^n K_n(N-x-1; p) \quad (14)$$

D. Tchebichef Polynomial

The weighted and normalized TP's n^{th} order can be defined as follows:

$$T_n(x) = \sqrt{\frac{\omega_T(x)}{\rho_T(n)}} (1-N)_n F_2(-n, -x, 1+n; 1-N; 1), n, x = 0,1,\dots,N-1 \quad (15)$$

This formulation involves the weight function of the TP, represented as $\omega_T(x)$, and the squared norm of the TP, described as $\rho_T(n)$.

$$\omega_T(x) = 1 \quad (16)$$

$$\rho_T(n) = (2n)! \binom{N+n}{2n+1} \quad (17)$$

These are the definitions of these quantities. $\frac{a!}{b!(a-b)!}, F_2$ specifies the binomial coefficients, while $3F_2$ represents the hyper-geometric function:

$$3F_2\,(-n,-x,1+n;1,1-N;1)\;=\;\sum_{k=0}^{\infty}\frac{(-n)_k\,(-x)_k(1+n)_k}{(1)_k(1-N)_k k!} \quad (18)$$

$$(a)_k = a(a+1)(a+2).(a+k+1) \quad (19)$$

The computation of the rising factorial $(a)k$ can be determined by employing the three-term recurrence relation. This approach is precious in mitigating numerical instability and computational overload that may arise from using hyper-geometric and gamma functions for calculating TP coefficient values.

$$T_n\,(x) = \beta_1\,T_{n-1}\,(x) + \beta_2 T_{n-1}(x)$$

$$\text{Where } n = 2,3,\dots,N-1, x = 0,1,\dots,N-1 \quad (20)$$

With initial conditions:

$$T_0\,(x) = 1/\sqrt{N} \quad (21)$$

E. Krawtchouk and Tchebichef (KT) Polynomials

When features in the provided domain meet requirements for localization and EC attributes, managing elements becomes significantly easier. These capabilities greatly impact the processing stages as the signal can be described by characterizing its contents using a small number of moments. EC will reduce computational complexity because fewer moments must be computed to represent the signal fully. By identifying the region of interest (ROI), localization attribute in space provides further computational savings and facilitates feature categorization:

$$\mathbb{R}_n\,(x;N) = \sum_{j=0}^{N-1}\chi_j\,(x;p;N)\dot{y}_j(n;p,N) \quad (22)$$

$$n,x = 0,1,\dots,N-1 \quad (23)$$

To put it mathematically, the polynomial obtained by merging two OPs is also orthogonal. Combining two fundamental orthogonal polynomials, $Y_n(x;\,p,N)\;and\;X_n(x;\,p,N)$, yields the expression for the hybrid polynomial form $\mathbb{R}_n(x)$ at n^{th} order. The parameters defining these orthogonal polynomials denoted by p and N are built utilizing a particular combination level. The following are the formulas for $X_n(x;\,p,N)\;and\;Y_n(x;\,p,N)$.

$$\chi_n(x;p,N) = \sum_{i=0}^{N-1}K_i(n;p)T_j(x) \quad (24)$$

$$y_n(x;p,N) = \sum_{i=0}^{N-1}K_i(x;p)T_j(n) \quad (25)$$

$$n,x = 0,1,\dots,N-1;\; p\in(0,1)$$

The equation below can be used to express the suggested hybrid OP as follows:

$$R_n\,(x;p,N) = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\sum_{l=0}^{N-1}K_j(i,p)T_j(x)K_l(n;p)T_l(i) \quad (26)$$

It is possible to express the equations using matrix notation. The polynomial forms of $R_n(x), X_n(x;\,p,N)$ $, Y_n(x;\,p,N), K_n(x), and\;T_n(x)$, respectively, correspond to the matrices $R, X, Y, R_k, and\;R_t$. In the equations below, it is demonstrated that the matrices KP and its transposition denoted as $R_k \equiv R_k^T$ are comparable. This significant observation highlights the symmetrical relationship between the KP coefficients along the major diagonal. Therefore, the related equations take on a particular form:

$$R = Y^T X \quad (27)$$

$$X = R_k^T R_t \quad (28)$$

$$Y = R_t^T R_k \quad (29)$$

$$R = (R_k R_t)^2 \quad (30)$$

In this context, the matrix transpose operator is denoted by $(\cdot)$. It is evident that $R_n(x)$ represents the squared expression of the sum of the Krawtchouk–Tchebichef polynomial (KP) and the Tchebichef polynomial (TP). Since the suggested set has a mathematical link with KP and TP, it can be represented by the squared Krawtchouk–Tchebichef polynomial (SKTP). Fig. 4 shows how the TP and KP modules generate the coefficients. It is noteworthy that, concerning KTP and TKP, SKTP has a notable localization characteristic and a higher EC property. Observing the SKTP coefficients symmetrically along the imaginary axis $x = (N-1)/2$ is possible. The SKTP coefficients of order $n = 127, 95, and\;64$ and those of order $n = 0, 31, and\;63$ are implied to correspond to this symmetry. In particular, $x = (N\,1)/2$ (imaginary axis), the SKTP coefficients range display symmetry are typically $n = \frac{N}{2}, \frac{N}{2}+1,\dots,N-1$. The values of $n = 0,1,\dots,N/2-1$ are included in these intervals. Moreover, the left half of the signal is associated with polynomial orders' coefficient within $n = 0,1,\dots,N/2-1$ range and the right half is related to the coefficients of polynomial orders within the range of $n = N/2, N/2+1,\dots,N-1$. Unlike TKP and KTP, where the indices of moments and signals are negatively correlated, the index of moments is linked to the indices of the signal in the signal domain. Furthermore, the SKTP basis functions for an $8*8$ block with diverse parameter p-values are shown in Fig. 3. The study reveals a gradual increase in low-frequency centre basis functions in vertical and horizontal orientations. The low frequency moves to the top left corner at a probability value of $p = 0.25$, as shown. Conversely, Fig. 3 illustrates that at a probability value of $p = 0.75$, the low frequency is displaced towards the lower right corner.

F. Regression-based Support Vector Machine (RSVM)

Support vector regression (SVR) was developed based on the support vector machine (SVM) methodology, primarily used for binary response variables. The fundamental concept behind SVR is constructing a tube with a width of ε around the data points, utilizing only residuals with absolute values more petite than a predetermined constant (referred to as ε-sensitivity). This concept is visually depicted in Fig. 5. In binary classification, two sets of points are established: those within a designated tube are not subject to penalties due to their proximity (within a predetermined threshold ε) to the predicted function. In contrast, those outside the tube are penalized based on their distance from the expected function. Support vector machines (SVMs) use a classification penalization strategy similar to this one. The methods of support vector machine learning (SVM) and support vector regression (SVR) rely on finding a hyper-plane that fits well in a feature space created by a kernel and can be efficiently generalized while maintaining the original features. Due to the scope of this book, an extensive examination of Support Vector

Regression (SVR) theory is not included. Whether SVR outperforms all other types of regression machines for continuous outcome prediction is also debatable. Because of this, we will now demonstrate how SVR is implemented in the e1071 library.

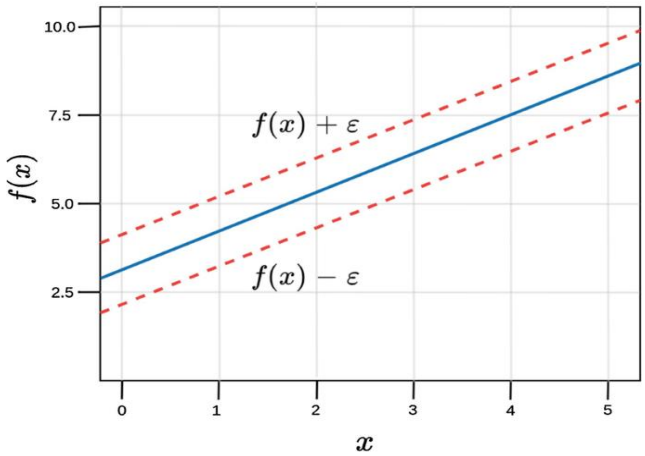


Fig. 5. Hyper-plane.

This part uses an FR system to evaluate the hybrid OP (SKTP) presented in terms of feature extraction. R-Support vector machines (SVM) are used in the facial recognition (FR) system's implementation. Fig. 5 shows the diagrammatic representation of the proposed method. The performance of KTP and TKP is evaluated against the hybrid OP, which can be generated using clear and noisy facial photographs within the FR system. The feature extraction process is the initial stage in the FR system. The image's height and width are extracted before any features are extracted using OP. We next use the given parameters, p and moment order, to generate the two polynomials, one for each dimension. Subsequently, the facial image is converted into the polynomial moment domain by applying the appropriate mathematical formula.

$$\Phi = R_1 \; if \; R_2^T \tag{31}$$

Here, R_1 and R_2 represent the produced polynomials and the face image. It is essential to highlight that the polynomial moment order determines the selection of the polynomial order. One can think of the generated moments (Φ) as a feature vector. The label of the matching face image ID is then appended to the feature vector. An array is created by combining all facial picture data in the database. The array is then normalized between -1 and 1 to guarantee that the features retain a constant dynamic range. SKTP and its hybrid versions, KP and TP, are evaluated for effectiveness and comparative analysis utilizing the AT&T facial database ORL. The database contains the faces of forty distinct persons. Each face image is 112×92 pixels in size. Ten photos of each person's face are taken at different times, with different lighting, expressions, and facial details. The RSVM model is trained using half of the face images during the training and testing phases. The moments taken from the data were used to train the RSVM model. The LIB-SVM was utilized to classify the data. RBF is employed in the RSVM model. Throughout the training phase, five-fold cross-validation was used to establish the gamma, cost and RSVM parameters. However, throughout the testing, various kinds of noise were added to every face image, making it challenging to assess SKTP's reliability. Three distinct forms of noise were employed during the degradation process: blur, salt and pepper, and Gaussian noise. Images without imperfections were utilized during the training phase exclusively. The classification accuracy is tested in the following ways to evaluate the classification's performance:

$$Accuracy = \frac{correctly\ predicted\ classes}{total\ testing\ classes} * 100\% \tag{32}$$

Ten separate testing and training images of faces were used for the testing and training stages to examine the stability of the categorization. The standard deviation and accuracy mean were computed for every ten iterations to evaluate the model's performance. Fig. 6 shows the separate training and testing phases in the face recognition system in this work.

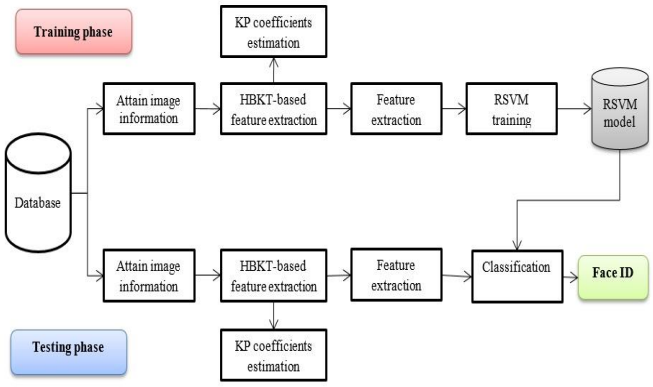


Fig. 6. Training and testing phases of the face recognition system.

G. Authentication

The evaluation of several continuous authentication systems presented in earlier research has yet to use various design criteria. A few studies consider limited requirements, such as accuracy and cost. Still, there needs to be a set of requirements for design that a continuous authentication system has to meet.

1) Passiveness: A system that is easy to operate shouldn't need constant user intervention. For instance, the user may find it very frustrating if the authentication system frequently requests that they re-enter their credentials.

2) Availability: A reliable authentication mechanism should always be available via the system. One major disadvantage of many previously suggested continuous authentication systems is that they are only sometimes available; they frequently fail because of insufficient data. In certain instances, a keyboard-centric system may erroneously decline access to a user engaged in movie viewing and not actively utilizing the keyboard.

3) High accuracy: High accuracy is, without question, the most crucial criterion for any authentication system. When a user attempts to log in as someone else, the system ought to be able to reliably and accurately identify them as imposters and deny their requests.

4) Scalability: With more users, the system should be able to manage an increasing volume of work. Its complexity in

terms of time and space should progressively increase with the number of users.

5) Efficiency: For two reasons, it is highly preferred to have a low response time or the amount of time needed to gather a test sample, analyze it, and render a conclusion. To promote user convenience, it is essential that the system rejects imposters and immediately authenticates legitimate users. Second, if there is a noticeable delay, security can also suffer. In the scenario where the authorization process requires five minutes, there is a potential vulnerability for an unauthorized individual to exploit the system and gain access to restricted resources during the on-going processing period, which lasts five minutes.

6) Low cost: The cost is a significant consideration regarding authentication methods used in low-security settings, like home computers. An authentication system should ideally be inexpensive to implement or alter in such contexts. Thus, systems without extra peripherals, like retina scanners, would often be preferred. However, expensive authentication systems could be used in highly secure areas like military bases.

7) Stability: Ideally, a feature should only slightly vary or retain its pattern for a specific time before being recorded for processing to facilitate authentication.

8) Extensibility: It should be possible for the authentication system to work on a broad range of devices independent of the underlying technology. No specific hardware should be needed for the system. One benefit of password-based authentication is its extensibility, which protects numerous methods, devices, and resources with limited system modification.

V. Results and Discussion

The proposed approach of remote user authentication based on face recognition is implemented in Python. The proposed work's environment is defined in Table I for when the face recognition based remote biometric user authentication system is implemented in real time on a login server.

After that, this work reviews five metrics employed to evaluate the suggested authentication system's accuracy. Traditionally, authentication systems are examined using the first three. To explore the precision of continuous authentication, we introduce two additional variables.

A. Metrics

The following metrics will be used in analyzing the results of applying the proposed method for face detection to perform facial user authentication in a remote biometric user authentication system:

1) False acceptance rate (FAR): The percentage of illegal users who are inadvertently approved in continuous authentication as opposed to the total number of fraudulent login attempts is known as the "False Acceptance Rate" (FAR). We use the notation $FAR_t = T_{EW}$ to describe the FAR below a given threshold, TEW. When security is the primary priority, it is advised to have a lower FAR.

2) False Rejection Rate (FRR): This indicator shows the percentage of legitimate requests mistakenly denied by users who have been granted authorized access to the system. The false rejection rate (FRR), when TEW is met, is shown by the notation $FRR_t = T_{EW}$. Because it improves user convenience, a lower FRR is preferred.

3) Rejection Rate (RR): This typically represents the total number of rejections (both legitimate and illegitimate) over the total number of attempts. In the context of authentication systems, this would mean how often access attempts are denied, regardless of whether those attempts are valid or invalid. RR Relationship to FRR (False Rejection Rate): The FRR is a subset of the RR, specifically focusing on the rate at which legitimate attempts (i.e., from authorized or genuine users) are incorrectly rejected. In contrast, RR could include all rejections, making it a broader measure. While FRR is a measure of system accuracy in identifying valid users, RR provides an overall view of the system's strictness or leniency in granting access.

4) Equal Error Rate (EER): Currently, it is noted that FAR and FRR are equivalent. However, because there is a trade-off between the two measurements, reporting FAR or FRR alone does not provide a complete picture. One rate can be lowered while permitting the other to rise. Therefore, we use EER (Equal Error Rate) rather than FAR or FRR to appropriately report the accuracy of H-RSVM (Context-Aware Behavior Analysis). This work uses $EER_t = T_{EW}$ notation to represent EER under TEW (Threshold Equal Weighting).

TABLE I. Recommended Login Server Hardware and Software Requirement

Requirements		
Sl. No	**Type**	**Specifications**
1	Processor	Intel(R) Core(TM)2 Duo CPU E8400 @ 3.00GHz 2.99 GHz processor (minimum hardware requirement)
2	Storage	195 GB (minimum storage requirement)
3	RAM	8.00 GB (minimum RAM requirement)
4	Display	Monitor resolution 1024x768 (minimum resolution requirement)
		Software Requirements
Sl. No	**Software type**	**Specification**
1	OS	Windows 10 Pro (or later)
2	S/W	Python (version 3.0 or later)

5) False Acceptance Worst-case interval (FAW): Within a specific time-frame, T, the authentication system generates accept/reject decisions as its output. Two potential output patterns can occur when an impostor attempts to authenticate in a ten-minute authentication period. Both sequences exhibit an equal number of incorrect authorizations. However, the second sequence is deemed more problematic as the impostor can exploit the system for a continuous duration of four minutes without being detected. It is a decline from the first sequence, in which the impostor was limited to one minute of system use. The most significant amount of time (minutes) that a fraudulent user could be mistakenly identified as authentic is what we refer to as FAW. FAW equals one minute in the first case and four minutes in the second.

6) False Rejection Worst-case interval (FRW): In a similar manner to the concept of a False Acceptance Window (FAW), the notion of a False Rejection Window (FRW) is introduced, denoting the maximum duration (measured in minutes) during which an authorized user may be mistakenly rejected and misidentified as an impostor.

7) Scalability metrics: The complexity of the authentication system, in terms of space and time, will experience a gradual increase as the number of users grows, as previously discussed. We use the well-known O notation to show the space and time complexity of the H-RSVM technique as a function of N, the number of individuals in the dataset, to measure the scalability of the suggested method.

We used Python to create an H-RSVM prototype to examine the authentication system's accuracy. The correctness of the model is usually assessed with a different data point set than those used to construct it. Thus, the dataset is divided into training and test sets to aid creation and assessment. However, evaluating the efficacy of a system processing time series data is unsuitable for the conventional K-fold cross-validation approach. Time series data comprises a series of measurements that frequently show local relationships between observations. Cross-validation ignores the structural characteristics introduced in the data by these dependencies. Therefore, in this work, we devised many experimental scenarios to assess the authentication system's accuracy instead of utilizing conventional cross-validation. These circumstances are discussed next.

8) Baseline: The available dataset was split into two equal portions to create the baseline scenario, which resulted in $TEW = TRW = 7h$. The model was trained over the half dataset, which comprised the first seven hours for every subject, and was tested on the second half. All bio-streams ($n = 0,1,2,..$) were used for system testing and training. R-SVM-based classification technique is used. Both linear and Radial Basis Function (RBF) kernels were used in the SVM example. Our experimental results show that the testing error diminishes and approaches the minimum while the training error decreases and approaches zero as the number of repetitions grows. As a result, we have chosen to run 40 iterations across each classifier. This work has provided the $EER_t = 7h$ value for each classifier.

9) Biased FAR_t/FRR_t: While it is simple to evaluate authentication methods using EER_t, in situations where high security is required, we may need to restrict F RRt to enhance user convenience or minimize FAR_t to prevent impostors from gaining access. A low FAR_t indicates strong security, while a low FRR_t ensures user convenience. We employ the same settings as the baseline in this experimental configuration. It is important to note that false rejection and false acceptance have different consequences. We analyze two situations: (i) measuring FRR_t and attempting to minimize FAR_t and (ii) measuring FAR_t and attempting to minimize FRR_t. The results for these two scenarios are presented in Table II. Despite a higher FRR, Table II shows that H-RSVM prevents acceptance of impostors. Table II demonstrates that H-RSVM does not negatively impact user convenience, as it correctly rejects impostors over 90% of the time without mistakenly leaving legitimate users.

10) Variable window size: This work set the training and testing window sizes to seven hours over the baseline. In this instance, we modify the testing and training window lengths as the total of the two amounts equals fourteen hours, with the training window length ranging from two to twelve hours. The average equal error rate (EER_t) for several classifiers relative to the training window size is also illustrated. When we increase the training window size from two to six hours, EER_t significantly decreases for all classifiers. After that, it stays the same until TRW approaches 11h. There are two possible reasons why EER has started to increase over this TRW. First, an over-fitting of the model is possible. Secondly, there could not be enough test points.

11) Moving training window: We tested various values for TEW and TRW. Our experiments revealed that this verification approach achieved the most favourable outcome when TEW and TRW were set to 4 hours. With 15 nodes' tree size, the classification method used was R-SVM, and the average equal error rate (EER_t) was determined to be 1.9%. The trained model must remain valid for the successive four hours to achieve maximum accuracy for TRW.

Table II provides the false acceptance rate (FAR), equal error rate (EER), accuracy, precision, recall, F-measure, and false rejection rate (FRR) metrics for the proposed RSVM method as compared to other state-of-the-art methods for the three different datasets.

Fig. 7 shows the performance comparison of the metrics such as accuracy, precision, recall, and F-measure for the proposed H-RSVM method compared with the other state-of-the-art methods when the ORL dataset is used for model training.

Fig. 8 shows the performance comparison of the metrics such as accuracy, precision, recall, and F-measure for the proposed H-RSVM method compared with the other state-of-the-art methods when the GT dataset is used for model training.

TABLE II. COMPARISON OF PROPOSED MODEL WITH EXISTING APPROACHES

Dataset	Method	FAR	EER	Accuracy	Precision	Recall	F-Measure	FRR
ORL Dataset	RSVM	0.00094	0.0199	98.82%	98.23%	98.19%	98.21%	0.030
	KNN	0.0009	0.045	95.80%	95.89%	92.7%	94.3%	0.389
	DT	0.018	0.210	94.6%	92.8%	94.5%	93.6%	0.389
	RF	0.020	0.035	97.9%	96.8%	94.2%	94.4%	0.333
	CNN	0.052	0.032	98.8%	97.9%	96.3%	95.5%	0.
GT Dataset	RSVM	0.0014	0.0360	96.73%	96.16%	96.08%	96.12%	0.0692
	KNN	0.0014	0.055	96.0%	94.9%	92.8%	91.7%	0.356
	DT	0.019	0.220	93.6%	91.8%	91.7%	91.2%	0.399
	RF	0.021	0.045	96.9%	95.0%	92.4%	93.0%	0.393
	CNN	0.053	0.032	95.8%	95.8%	94.3%	93.5%	0.600
FACE94	RSVM	0.0005	0.0397	95.9%	95.83%	95.16%	95.49%	0.0784
	KNN	0.0005	0.045	95.0%	93.9%	91.7%	90.2%	0.349
	DT	0.021	0.210	92.6%	91.6%	90.7%	89.0%	0.389
	RF	0.023	0.035	95.9%	94.4%	92.4%	88.4%	0.833
	CNN	0.024	0.032	94.8%	93.9%	93.5%	92.0%	

TABLE III. RR, AUTHENTICATION DELAY, AND AUC COMPARISON WITH EXISTING APPROACHES

Rejection Rate (RR)	**ORL**	**GT**	**FACE94**
Proposed	0.9719	0.9407	0.9315
TPTSSR	0.9333	0.9233	0.9133
LPP	0.9357	0.9257	0.9057
CSDL-SRC	0.955	0.935	0.945
Deep CNN	0.91	0.901	0.900
SIFT	0.955	0.945	0.925
Authentication Delay (ms)	**ORL**	**GT**	**FACE94**
Proposed	1.3568	1.5405	1.4767
HMM-UBM	8	9	10
Conv-DCWRNN	15	16	17
AUC	**ORL**	**GT**	**FACE94**
Proposed	0.9804	0.97	0.95
DT	0.97	0.96	0.94
RF	0.92	0.91	0.90

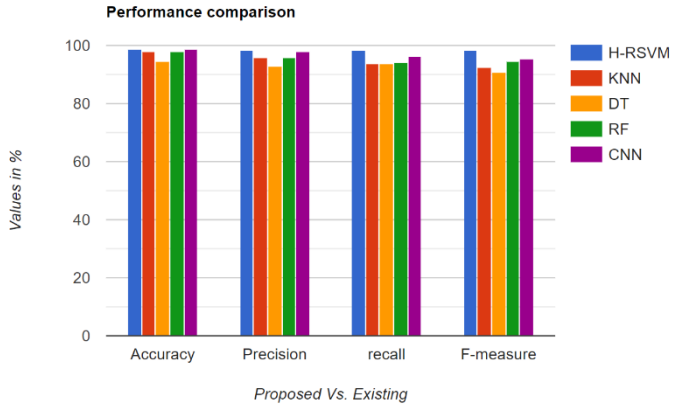


Fig. 7. Performance comparison with the ORL dataset.

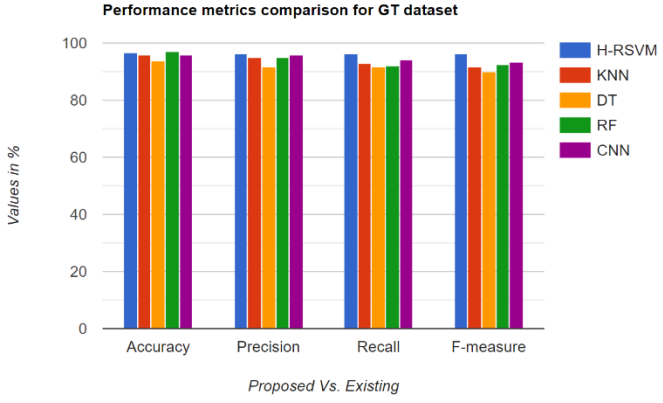


Fig. 8. Performance comparison with the GT dataset.

Fig. 9 shows the performance comparison of the metrics such as accuracy, precision, recall, and F-measure for the proposed H-RSVM method compared with the other state-of-the-art methods when the FACE94 dataset is used for model training.

Fig. 10 shows the performance comparison of other metrics such as FAR, EER, and FRR for the proposed H-RSVM method compared with the other state-of-the-art methods when the ORL dataset is used for model training.

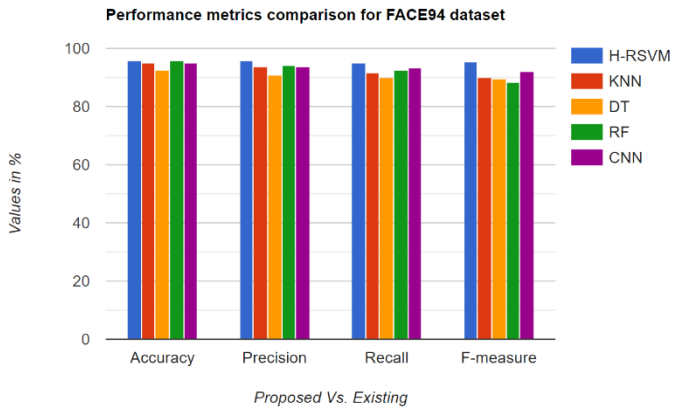


Fig. 9. Performance comparison with FACE94 dataset.

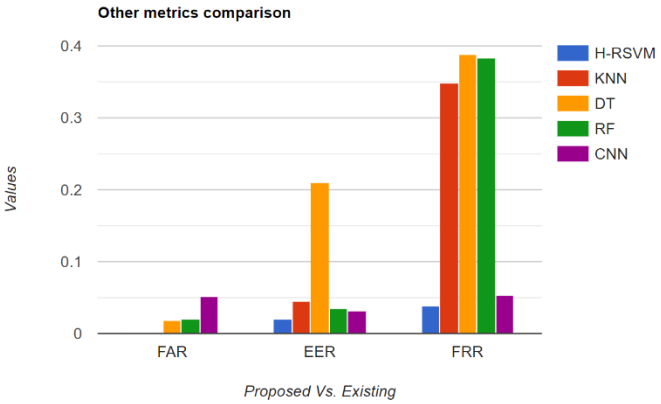


Fig. 10. Other metrics comparison with ORL dataset.

Fig. 11 shows the performance comparison of other metrics such as FAR, EER, and FRR for the proposed H-RSVM method compared with the other state-of-the-art methods when the GT dataset is used for model training.

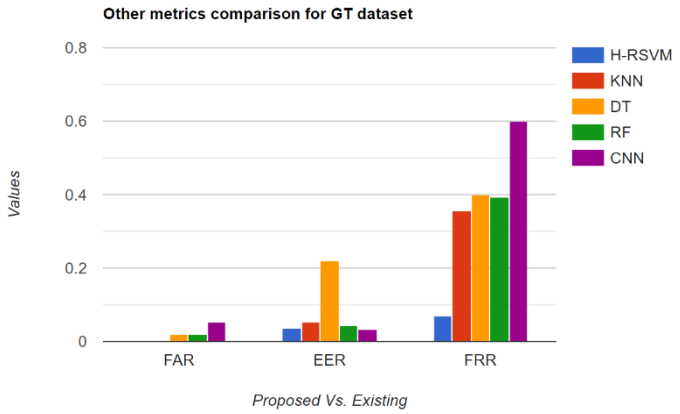


Fig. 11. Other metrics comparison with the GT dataset.

Fig. 12 shows the performance comparison of other metrics such as FAR, EER, and FRR for the proposed H-RSVM method compared with the other state-of-the-art methods when the FACE94 dataset is used for model training.

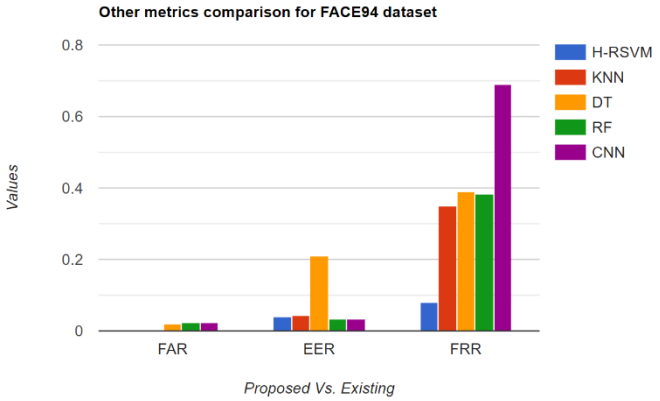


Fig. 12. Other metrics comparison with the FACE94 dataset.

Fig. 13 shows the performance comparison of another metric RR for the proposed H-RSVM method compared with other state-of-the-art methods for all datasets used in training.

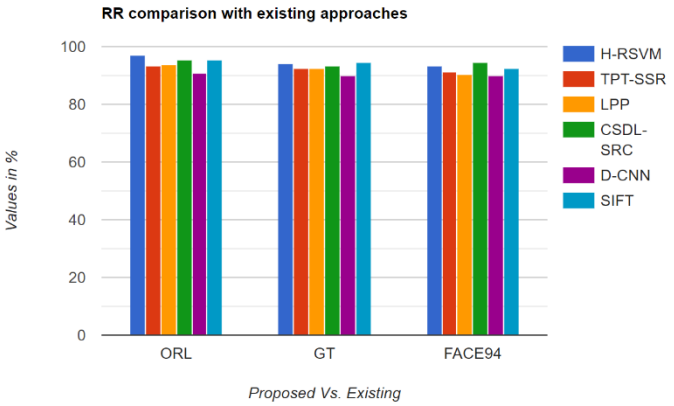


Fig. 13. RR comparison with existing approaches.

Fig. 14 shows the performance comparison of another metric, authentication delay, for the proposed H-RSVM method compared with the other state-of-the-art methods for all three datasets used for model training.

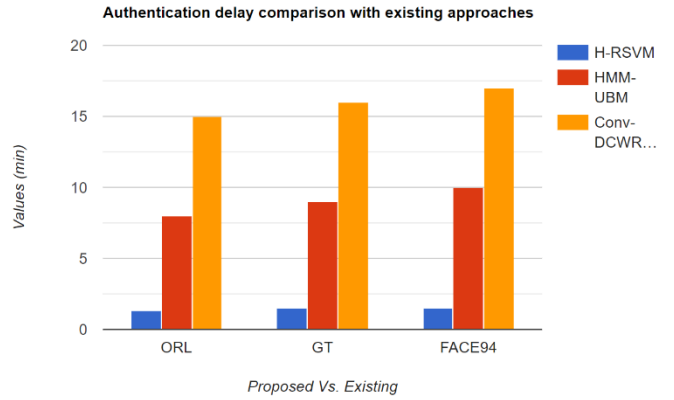


Fig. 14. Authentication delay comparison with existing approaches.

Fig. 15 shows the performance comparison of another metric, area under curve (AUC), for the proposed H-RSVM method compared with other state-of-the-art methods for all three datasets used for model training.

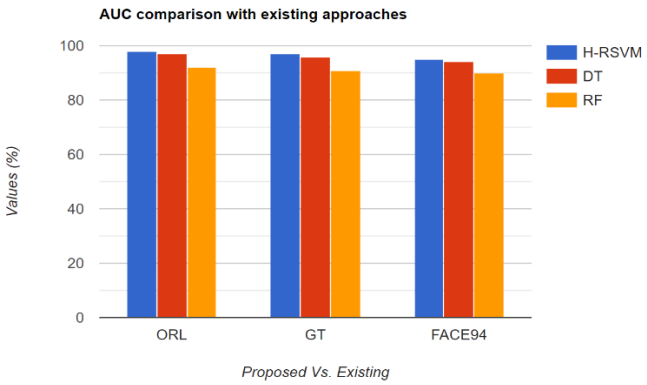


Fig. 15. AUC comparison with existing approaches.

Fig. 16 and Fig. 17 depict failed authentication and successful authentication use cases, respectively, using the proposed face-recognition-based remote authentication system. The model gives the correct prediction outcomes for both non-authorized user and authorized user with high rate of prediction accuracy. Based on the experimentation, it is proven that the anticipated model works well compared to other approaches, as discussed further below.

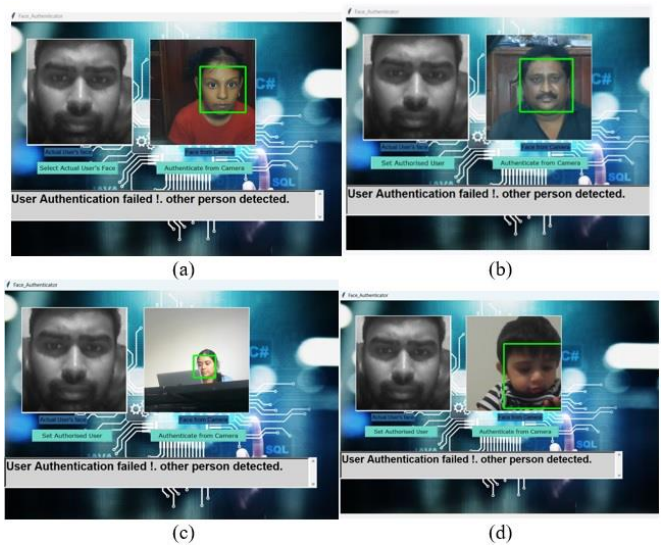


Fig. 16. Authentication outcomes for non-authorized users (a-d).

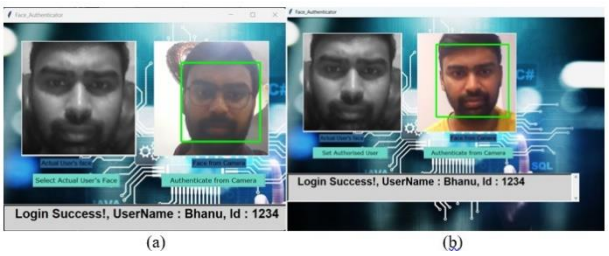


Fig. 17. Authentication outcomes for an authorized user (a-b).

Table II compares existing approaches like KNN, DT, RF and CNN. Metrics like FAR, EER, FRR, accuracy, precision, recall and F-measure are evaluated and compared. The proposed model gives 98.82% accuracy for the ORL dataset, 96.73% accuracy for the GT dataset, and 95.9% accuracy for the FACE94 dataset. The proposed model offers 98.23%, 96.16%, and 95.83% precision for the ORL, GT, and FACE94 datasets. It gives 98.19%, 96.08%, and 95.16% recall for ORL, GT, and FACE94 datasets. The proposed H-RSVM model gives 98.21%, 96.12%, and 95.49% F-measure for ORL, GT, and FACE94 datasets (see Fig. 7-12). Table III compares RR, authentication delay, and AUC evaluation with other approaches. The proposed model gives 98.04% AUC for ORL, 97% for GT, and 95% for the FACE94 dataset. The authentication delay is lesser for the proposed model, and the RR of the anticipated model is 97.19% for ORL, 94.07% for GT, and 93.15% for the FACE94 dataset, respectively (see Figures 13-15). Based on the observations, it is noted that the proposed model gives satisfying outcomes compared to other approaches.

B. Real-time Authentication

We define authorization at the outset of this section. Next, this work presents a real-time adaptive authorization (RAA) system that leverages H-RSVM's judgments to offer a tremendously flexible access control strategy. The RAA concept is not exclusive to H-RSVM. It provides a flexible model of access control that can be implemented in any authorization system that allows users to access content based on choices made by a continuous authentication method. Authorization determines whether to grant access to a system, resource, or location to a person who has already undergone authentication. Conventional authorization techniques use the user ID of an authenticated user to assign them a specific access level. However, as demonstrated by the non-zero FRR of continuous authentication systems, this simple method may unintentionally block allowed access if the authentication system cannot identify a legitimate user for a short time. Imagine a situation where a personal laptop is secured against unauthorized users using a continuous authentication method. The authentication technique authenticates the user at first. Using user ID, the authorization system establishes the user's access level. Nonetheless, the laptop can log users out if the authentication process unintentionally rejects them. The inconvenience resulting from false rejections can be reduced by using RAA techniques. These techniques continually adjust the user's access level following the authentication system's most recent determination. Here, we introduce an RAA technique with an ongoing authentication scheme.

A trust-based RAA modifies the user's access level in real-time using the trust level (TRL) parameter. Based on past choices made by the continuous authentication system, the recently established TRL metric measures our degree of confidence in a user. A higher number on the TRL scale, which ranges from 0 to 100, denotes a higher degree of trust. TRL has an initial value of 100 when a user is approved and authenticated. Following every user authentication, the trust update method continuously updates the TRL value. A simple technique for updating trust could involve increasing or decreasing TRL by a fixed amount following every accept or reject choice. The pseudo-code for this method is displayed in the trust update procedure. Two parameters, W_{accept} *and* W_{reject}, need to be set. When we are positive that a user is authentic, we should set the TRL value to 100. Otherwise, it has to be set to 0 immediately as we learn that an impostor exists. It indicates that the values of

W_{accept} *and* W_{reject} need to be chosen appropriately. In the worst-case situation, the authentication system can inadvertently deny three consecutive requests from legitimate users if, for instance, an SVM classifier with 15 nodes' tree size yields a false rejection rate of 3. The RAA approach concludes that a user is fraudulent ($TRL = 0$) if the authentication system rejects a request four times in a row. With this classifier, we can use the following formula to set the value of W_{reject}: $-100\ FRW + 1 = -100\ 4 = -25\ is\ W_{reject}.FAW = 4$ means that, in the worst situation, four successive tries would lead to mistakenly identifying an impostor as a valid user. TRL should be more prominent than 100 if the authentication system accepts five requests in a row. Therefore, the formula for setting W_{accept} is as follows: $W_{accept} = +100\ FAW + 1 = +100\ 5 = +20$. Various purposes may require different threshold levels to be specified. The work decided to restrict the threshold value for these accounts to 100 to ensure that the user can evaluate input images when the system is positive they are real. However, a lesser degree of trust might be adequate for less sensitive activities, like basic web browsing. When combined with RAA, H-RSVM can improve user-friendliness without compromising the high-security standards required for critical applications.

VI. Conclusion and Future Work

This work develops a continuous remote biometric user authentication system based on a face recognition model pre-trained on face image datasets. This research work implements an algorithm combining HB-KT Polynomials and RSVM for a face recognition based remote user authentication system that implements a model pre-trained on the ORL, Face94 and GT datasets to perform face recognition for continuous remote biometric user authentication. Comprehensive analysis has been performed to ascertain the accuracy and performance of the proposed method as compared to previous state-of-the-art algorithms. A functioning prototype has also been implemented as a Python program. The proposed model gives accuracies of 98.82% (ORL dataset), 96.73% (GT dataset), and 95.9% (Face94 dataset). Compared to other previous state-of-the-art models, the proposed model generally has been observed to have higher F-measure, accuracy, and speed. More research should be made on the use of multimodal biometric data, taken from multiple channels including face recognition, voice, and behavioral characteristics to improve the performance of continuous user authentication. Developing adaptive machine learning algorithms tailored for user interfaces that can produce personalized methods based on the particular interaction can help consider the change in behavior depending on context or mood. Privacy related issues need to be resolved, which requires storage and processing to be secure and possible solutions include decentralized approaches or homomorphic encryption. Moreover, context aware computing can adapt the authentication requirements depending on the ongoing situation context factors, usability studies are vital for the assessment of users' acceptance and interaction with the established systems, ensures both secure and usability.

Acknowledgment

This research was performed at Texas A&M University-Kingsville in a project funded by the Department of Homeland Security (DHS) through the grant no. 21STSLA00011-01-00. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

References

[1] S. Saragih, J. Lucey, and J. Cohn, "Deformable model fitting by regularized landmark mean-shifts," Int. J. Comput. Vis., vol. 91, pp. 200–215, 2010.

[2] T. Sujatha, N.R. Wilfred Blessing, S. Anand, and E. Daniel, "Automated face authentication and recognition using deep neural network with SVM classifier in cloud environment," in *Disruptive Technologies for Big Data and Cloud Applications*, J.D. Peter, S.L. Fernandes, and A.H. Alavi, Eds., vol. 905, Springer, Singapore, 2022.

[3] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikainen, "Deep learning for generic object detection: a survey," *arXiv preprint arXiv:1809.02165*, 2018.

[4] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1492–1500, 2017.

[5] R.J. Wang, X. Li, and C.X. Ling, "Pelee: a real-time object detection system on mobile devices," in *Advances in Neural Information Processing Systems*, pp. 1963–1972, 2018.

[6] L. Li, X. Feng, Z. Xia, et al., "Face spoofing detection with local binary pattern network," *J. Vis. Commun. Image Represent.*, pp. 182–192, 2018.

[7] S. Ren, K. He, R. Girshick, et al., "Faster R-CNN: towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, pp. 91–99, 2015.

[8] J. Redmon, S. Divvala, R. Girshick, et al., "You only look once: unified real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 779–788, 2016.

[9] M. Ramgopal, M.S. Roopesh, M.V. Chowdary, et al., "Masked facial recognition in security systems using transfer learning," *SN Comput. Sci.*, vol. 4, p. 27, 2023.

[10] J. Imran and B. Raman, "Deep motion templates and extreme learning machine for sign language recognition," *Vis. Comput.*, 2019.

[11] S. Ravi, M. Suman, P.V.V. Kishore, K. Kumar, and A. Kumar, "Multi-modal spatio-temporal co-trained CNNs with single modal testing on RGB–D based sign language gesture recognition," *J. Comput. Lang.*, vol. 52, pp. 88–102, 2019.

[12] G. Hu, Y. Yang, D. Yi, J. Kittler, W. Christmas, S.Z. Li, and T. Hospedales, "When face recognition meets with deep learning: An evaluation of convolutional neural networks for face recognition," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Santiago, Chile, pp. 142–150, 2015.

[13] P.S. Prasad, R. Pathak, V.K. Gunjan, and H.V.R. Rao, *Deep Learning Based Representation for Face Recognition*. Springer, Berlin, Germany, pp. 419–424, 2019.

[14] Y. Liu, M. Lin, W. Huang, and J. Liang, "A physiognomy-based method for facial feature extraction and recognition," *J. Vis. Lang. Comput.*, vol. 43, pp. 103–109, 2017.

[15] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, pp. 1701–1708, 2014.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, pp. 1904–1916, 2015.

[17] J. Yu, K. Sun, F. Gao, and S. Zhu, "Face biometric quality assessment via light CNN," *Pattern Recognit. Lett.*, vol. 107, pp. 25–32, 2018.

[18] R. Singh and H. Om, "Newborn face recognition using deep convolutional neural network," *Multimed. Tools Appl.*, vol. 76, pp. 19005–19015, 2017.

[19] Y. Sun, X. Wang, and X. Tang, "Hybrid deep learning for computing face similarities," *Int. Conf. Comput. Vis.*, vol. 38, pp. 1997–2009, 2013.

[20] G.P. Nam, H. Choi, and J. Cho, "PSI-CNN: A pyramid-based scale-invariant CNN architecture for face recognition robust to various image resolutions," *Appl. Sci.*, vol. 8, p. 1561, 2018.

[21] D. Menotti, G. Chiachia, A. Pinto, W.R. Schwartz, H. Pedrini, A.X. Falcao, and A. Rocha, "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE Trans. Inf. Forensics Secur.*, vol. 10, pp. 864–879, 2015.

[22] C. Qin, X. Lu, P. Zhang, H. Xie, and W. Zeng, "Identity recognition based on face image," *J. Phys. Conf. Ser.*, vol. 1302, p. 032049, 2019.

[23] Z. Zhu, P. Luo, X. Wang, and X. Tang, "Recover canonical-view faces in the wild with deep neural networks," *arXiv preprint arXiv:1404.3543*, 2014.

[24] Z. Lu, X. Jiang, and A.C. Kot, "Deep coupled ResNet for low-resolution face recognition," *IEEE Signal Process. Lett.*, vol. 25, pp. 526–530, 2018.

[25] Y. Zhang, D. Zhao, J. Sun, G. Zou, and W. Li, "Adaptive convolutional neural network and its application in face recognition," *Neural Process. Lett.*, vol. 43, pp. 389–399, 2016.

[26] B. Abd El-Rahiem, M. Amin, A. Sedik, F.E. Abd El Samie, and A.M. Iliyasu, "An efficient multi-biometric cancellable biometric scheme based on deep fusion and deep dream," *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 4, pp. 2177–2189, 2022.

[27] A. Rengaraj, A.R. Kishan, A. Abraham, and A. Sattenapalli, "Centralized intelligent authentication system using deep learning with deep dream image algorithm," in *Advances in Power Systems and Energy Management*, Springer, Singapore, pp. 169–178, 2021.

[28] E.A. Elshazly, F.G. Hashad, A. Sedik, and N. Abdel-Salam, "Compression-based cancelable multi-biometric system," *Research Square*, Nov. 2022.

[29] D. Lu and L. Yan, "Face detection and recognition algorithm in digital image based on computer vision sensor," *J. Sensors*, vol. 2021, pp. 1–16, Sep. 2021.

[30] H.R. Vijaya Kumar and M. Mathivanan, "A novel hybrid biometric software application for facial recognition considering uncontrollable environmental conditions," *Healthcare Anal.*, vol. 3, 2023.

[31] S. Hangaragi, T. Singh, and N. Neelima, "Face detection and recognition using face mesh and deep neural network," *Procedia Comput. Sci.*, vol. 218, pp. 741–749, 2023.

[32] R. Sharma and A. Ross, "Periocular biometrics and its relevance to partially masked faces: A survey," *Comput. Vis. Image Underst.*, vol. 226, 2023.

[33] G. Rajeshkumar, M. Braveen, R. Venkatesh, P. Josephine Shermila, B. Ganesh Prabu, B. Veerasamy, B. Bharathi, and A. Jeyam, "Smart office automation via faster R-CNN based face recognition and internet of things," *Measurement: Sensors*, vol. 27, p. 100719, 2023.

[34] J. Mason, R. Dave, P. Chatterjee, I. Graham-Allen, A. Esterline, and K. Roy, "An investigation of biometric authentication in the healthcare environment," *Array*, vol. 8, 100042, 2020.

[35] A. Saini, "Analysis of different face recognition algorithms," *Int. J. Eng. Res. Technol.*, vol. 3, no. 11, pp. 235–239, 2014.

[36] "Cambridge ORL Face Database," available at: https://cam-orl.co.uk/facedatabase.html.

[37] "AT&T Lab. Cambridge Face Database," available at: www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

Designing Conversational Agents for Student Wellbeing

An Exploratory Study of User Acceptance and Expectations

Jieyu Wang[1], Li Zhang[2], Dingfang Kang[3], Katherina G. Pattit[4]*
Herberger Business School, St. Cloud State University, St. Cloud, United States of America[1, 2, 3]
Office of the Provost, St. Cloud State University, St. Cloud, United States of America[4]

***Abstract*—The innovative development of AI technology provides new possibilities and new solutions to the problems that we are facing in modern society. Student well-being status has been a major concern in well-being care, especially in the post-pandemic era. The availability and quality of well-being support have limited the accessibility of well-being resources to students. Using conversational agents (CA) or chatbots to empower student well-being care is a promising solution for universities, considering the availability and cost of implementation. This research aims to explore how CAs assist students with possible well-being concerns. We invited 96 participants to fill out surveys with their demographic information and 11 short answer questions concerning their well-being and their acceptance and expectations of CAs. The results suggested the participants accepted the use of well-being CAs with ethical concerns. Upon user acceptance, the participants expressed expectations on design features such as facial expression recognition, translation, images, personalized long-term memory, etc. Based on the results this work presents a conceptual framework and chat flows for the design of a student well-being chatbot, which provides a user-centered design example for UX designers of wellbeing. Further research will introduce detailed design discoveries and a high-fidelity CA prototype to shed light on student well-being support applications. Implementing the CA will enhance the accessibility and quality of student well-being services, fostering a healthier campus environment.**

***Keywords*—Conversational agent/chatbot; wellbeing; UX design**

I. Introduction

The world is advancing with the fast development of AI chatbots. For students in schools and institutes, the expectation and stress upon them can be astounding. Student well-being has a tremendous impact on their performance in their study, work, and personal lives. Unavoidable life mishaps may cause a variety of well-being problems that require external interference. Schools have traditional ways to support their students in their physical health. However, assistance to student well-being has always been limited by the accessibility and availability of psychological and counseling resources.

Nowadays, we have more advanced artificial intelligence that empowers the work in industry and academia. Conversational agents (CAs) or chatbots, for their high availability and capability to interact with humans in natural language communication, would be a promising solution to address the shortage of resources for student well-being care.

The primary objective of this study is to ensure student well-being through the utilization of a chatbot. To achieve this goal, we sent out surveys to collect students' feedback about their acceptance and anticipation of chatbots. Based on their responses, we have developed a student well-being chatbot prototype that is embedded with information on some of the common well-being issues that students usually encounter during their education. This paper presents the design framework and chat flows of this prototype. We aim to reduce the time and confusion of students to search for relevant information online, allowing them to retrieve the needed internal university information and external resources that can be suitable and helpful to their situations.

II. Related Works

Researchers have done comprehensive examinations of CA engagement in well-being. They were able to recognize the positive impact of CAs on major mental-health symptoms for CAs that are AI-based, multimodal, highly accessible, and externally connected. They also identified the key features of CAs that would shape user experience in well-being effectiveness and controlled integration [1]. The availability of well-being CA-embedded applications in the market from January 2020 to 2023 also has been reviewed. A total of 18 applications have been selected, classified, and analyzed. Most of the applications are aimed at multiple well-being issues and have cognitive behavioral therapy implemented. The proportion of AI-based apps and rule-based apps is equal with a total of 48 behavior change techniques utilized [2]. In this related work, we reviewed the studies of CAs for student well-being at universities, CAs for well-being outside universities, chatbot design with specific features, and ethical concerns.

A. Well-being Assistants at Universities

School environments have a significant impact on the construction of student wellness. The relationship construction between students and school has been a significant factor that impacts student well-being [3]. The persistent stress students endure during their education has long been identified as a major individual well-being issue. During the pandemic, the demand for mental health assistants increased. Nelekar et al. (2022) investigated the effectiveness of using an embedded conversational agent (ECA) to assist students with their well-being. They addressed the importance of cultural context, social environment, and personal intention in their design for the ECA.

The result demonstrated a positive effect of ECA on stress reduction and student construction [4].

Aiming at the same goal, Jeong et al. (2023) imported a robot to provide therapeutic intervention as well as social tips. Their result suggested that using AI products could be a promising solution for well-being care. Follow-up interviews and analysis revealed the significant factors that should be considered during the development of a well-being caring agent, such as personality traits, underlying behaviors, and association quality between the participant and the robot [5].

Using GPT as the base, a research team with expertise in psychology has developed a conversational agent that matches real-world context. It also included the response questions that were frequently asked by students. The team evaluated the CA's performance and participant's responses. The feedback favored the approach of using GPT for well-being care and suggested GPT would require personalized features to provide better services to users [6].

Usta's (2021) study addressed the harmful stress among many undergraduate students. To discover insight into the problem, the team conducted individual interviews with participants and arranged case studies. As identified by the analysis, the stress mainly originated from maintaining a highly productive status with self-obligation. Additionally, the team was able to summarize the primary needs of the student during the stress. The case study involved the utilization of a chatbot that prompted the participants with encouragement and suggestions to increase their self-esteem. Further research was advised to consider the identified needs in this study and a different setup on the case study for a more comprehensive outcome [7].

B. The CAs for Well-being Care

CAs are also applied in mental healthcare in the general environment. To address the problem that self-help resources are not able to last long enough, Ly et al. (2017) implemented the interventions' strategies, provided solutions using a chatbot, and published it as an application to facilitate well-being care. They invited 28 participants and divided them into two groups for examination and a comparison study. The analysis of follow-up data resulted in higher achievement when compared to other intervention-implemented chatbots in related fields. This study presented self-help intervention as an effective tool for well-being care support to public health [8].

Inkster et al. (2018) conducted research aiming to address the mental health resource shortage. Their study focused on the evaluation of an AI-embedded conversational application. They invited anonymous users from a global range as participants in this study and examined the influence of the app on the participants. Evaluations were based on user feedback to evaluate the functionality of the app. Preliminary data facilitated the initial speculation. However, further research and analysis with a larger sample size were required to justify the conclusion [9].

Chatbots were also used to facilitate the well-being status of invited social groups. The participants were examined in 10 days and feedback was collected with follow-up interviews. The comparison of the experimental and controlled groups supported that the chatbot had a significant effect on improving the well-being status of the participants. Interaction with the chatbot was favored by the majority of the participants with positive feedback. Further development and research should involve the improvement of current functionality and personalization, as well as conducting a study with a larger sample size [10].

However, a study showed an insignificant impact of CA on improving students' mental well-being. An evaluation of a well-being care application has been conducted to support the increasing well-being care demand. The research team invited two groups of undergraduates and performed a study for each with randomized control. The outcome did not demonstrate the influence of the application on participants' well-being status. The insignificant result indicated the importance of appropriate CA design [11].

Aside from general well-being, some CA studies addressed specific well-being issues. Adolescence is considered a significant period in life that can easily lead to serious mental health issues if not handled properly. To support the mental health of adolescents, Gabrielli et al. (2020) introduced a chatbot as a possible resource. The study involved the design and evaluation of chatbot software that implemented CRI and CRIS coaching sessions. The result of the study favored this design with most of the participants providing positive feedback and recommendations. A follow-up investigation was required in the future to capture the effects of this design in the long-term period [12].

Because the effects of psychological interventions for cancer treatments are significant, chatbots are applied to help facilitate the process to make it more accessible and acknowledged by young adults. A study introduced the participants to a chatbot (Vivibot) to teach them psychological skills that could help them with cancer treatment and examined the effects in four weeks. The result showed a reduction in anxiety levels by using the chatbot and no other impactful effects on other well-being issues [13].

C. CA Design with Diverse Features

Emotionally sensitive chatbots have been under-explored. Ghandeharioun et al. (2019) conducted a study on an emotionally aware chatbot. They designed and evaluated the chatbot with 39 participants in two weeks. The result suggested that the participants supported the emotion detection feature. The participants' feedback has been collected and summarized into several guidelines that primarily focused on participants with positive moods. These guidelines were to be addressed in their further study [14].

Participatory design has been used to design a chatbot. A research team collected user reflections on the needs of a chatbot and summarized several guidelines for chatbot design that involved information usage and the delivery method for user-centered design. It involved the cooperation of both the evaluation and effort from the designer and mental health experts to implement these needs. Future studies would also consider possible ethical concerns and the practicality of this feature regarding the development of technologies [15].

Other than inviting users to participate in the design phase, feedback after the experiment can also be used to support CA

integration. Demirci (2018) used a conversational agent to manage interactions required for well-being care to the participants and explored the qualities of good design of CA for well-being care. They organized their results from user engagement to design qualities. They suggested that further research could repeat the process with a different user group or even a CA for a different purpose [16].

The development of conversational agents and smart speakers provides a possibility for new methods for well-being data collection. Maharjan et al. (2022) explored the feasibility of collecting self-report data with a speech-recognition CA. It presented a challenge to collect data with such a device, especially when deployed in a home context. They also conducted interviews and analyzed the data that reflected the technological limitations and appropriateness of the CA. This research provided valuable experience in using speech-recognition CA to support well-being care [17].

D. Ethical Issues

Ethical concerns in CA design are essential. Yet the appropriateness of applying them to well-being services is under-explored. Research has been done to raise awareness among researchers and developers on the advantages and downsides. A research team analyzed 120 posts regarding well-being support applications and identified the features that other researchers could pay attention to or should be avoided. The team called for the consideration of responsibility and social effects of the application [18].

Privacy and data security of users are one of the major concerns for well-being CAs. Gengoux and Roberts (2019) focused on the ethical use of user data when dealing with mental health problems in higher education [19]. Local policies are also a concern when introducing a CA to support well-being services in the context of universities. Graham et al. (2023) focused on the school's influence on the construction of wellness upon students regarding safety and internal well-being. The study was based on interviews with both the school and their students. The team identified the relationship between the school practice and well-being conditions by taking different aspects into the analysis process. These findings highlighted the significance of relationship construction of the school with students [20].

In this work, nearly one hundred university students who present common student well-being issues were recruited. They were encouraged to report well-being issues that needed to pay attention to in this research. By addressing these issues, we aim to gain a comprehensive understanding of student well-being problems and identify what assistance and resources we can provide. In the surveys, participants were asked questions to assess their perception of the necessity of a student well-being conversational agent and their expectations of the functions. The research questions of our study include the following four points:

- User well-being status: The participants were asked to identify their well-being status. This allowed them to review and pinpoint the causation related to their current well-being status, as well as areas that were presented to have a positive or negative influence on their problem.
- User acceptance: The participants were questioned about their opinions about using AI products such as CAs for student well-being support. They were encouraged to freely express their feelings toward this choice. They were welcome to express their support or disfavor with reasonable explanations. The feedback informs us where and how we can support students with our prototype.
- User expectations: The participants were asked to express their expectations on the use of conversational agents to assist them on wellbeing issues. They were encouraged to provide reasons for their responses and suggestions on specific tasks the CA would be capable of, which could be helpful and accessible to improve their well-being or support them.
- Suggestions for design and improvement: The participants were asked to provide suggestions for the prototype. This included specific features or functionalities to provide qualified services, significant issues to pay attention to and address, or user interface elements that could enhance user experience. Our team collected and evaluated these suggestions and eventually converged into several categories to provide the conceptual framework and guidelines for the design of the prototype.

By exploring the feedback from these questions, we aim to gather valuable insights from the participants to design and improve the prototype, addressing their needs and preferences to create a more user-friendly and valuable tool for student well-being support.

III. METHODOLOGY

A. Participants

The study recruited 96 students from a university located in the Midwest region of the United States. We collected and initially analyzed the demographic information to understand the participants. Our participants are students from the age of 19 up to 37 with an average of 23.34 (SD=3.35). The gender distribution contains 59% (57 participants) male and 41% (39 participants) females.

The participants were from nine diverse majors: Accounting, Economics, Education, Finance, General Business, Information Assurance, Information Systems, Management, and Marketing, Among the different majors, the participants were mainly distributed among Information Systems, Information Assurance, Finance, Marketing, and General Business. Table I contains the number of participants in different majors. Two participants did not disclose their majors.

TABLE I. PARTICIPANT BACKGROUND

MAJOR	PARTICIPANT NUMBER
Marketing	10
Accounting	6
General Business	12
Information systems	24
Management	4
Finance	15
Mathematics/Education	1
Economics	1
Information Assurance	21

Before their participation, the study received approval from the Institutional Review Board (IRB), and the students were invited to join the experiment after providing their informed consent by signing the consent forms.

B. Surveys and Data Analysis

Ninety-six participants participated in our survey studies. They provided answers to their demographic information and 11 short-answer questions concerning their well-being and their acceptance and preferences of chatbots. The collected data were analyzed using qualitative data analysis techniques that involve the open coding method (OCM). OCM allows us to uncover some of the hidden information within the participants' surveys. As we continued the process, we were able to identify the interrelationship between their answers and further categorize them into various themes.

The open coding was conducted by two researchers to identify preliminary concepts of interest. Due to the nature of the materials used, measuring inter-coder reliability was challenging. Throughout the analysis, the two researchers engaged in discussions to explore similarities and differences in the themes that emerged from the data. This iterative process allowed for a deeper understanding of the data and the development of meaningful interpretations.

IV. Results and Discussion

The study aimed to uncover insights and patterns within the collected data, facilitating a comprehensive understanding of the participants' experiences and perspectives related to the well-being CA and its usage in educational institutions.

Our open-coding analysis clarified the following design guidelines. We will center our design to address the major issues.

A. Methods to Maintain Well-being

The participants were questioned about the existence of any of the symptoms such as stress, anxiety, depression, and insomnia that they experience. The feedback from the survey indicated that the pandemic was one of the causes of the aforementioned symptoms. Further analysis of the feedback gave us a more comprehensive perspective on the problem.

Participant 4 suffered depression during the pandemic because of restricted socialization under self-quarantine. Participant 10 also experienced all the symptoms during the pandemic, especially when they were working in a hospital.

"I suffered from depression when I was in the Pandemic. The reason is that I had to self-quarantine for two weeks whenever I went back to Korea, but I once suffered from depression because I couldn't go out and meet people." (Participant 4)

"Yes I have experienced all mainly from the harsh transition to online schooling and not being able to go out and socialize as much during the pandemic and I had also worked at the Hospital during the pandemic which was stressful." (Participant 10)

Other participants showed effective self-coping ability, and socialization helped them stay mentally healthy.

Participants 34 and 57 were in a completely different situation in contrast to the previous examples. Participant 34 was not subjected to any of the symptoms and was able to keep up the good status. Participant 57 didn't experience any of the symptoms and had dependable communication channels with others. Their need for an additional release channel was much less than many other participants and less likely to suffer some well-being issues.

"Explicitly No because I tried to familiarize myself with digital transformation and got some new skills. Those skills were very helpful to utilize to get ease in my routine life. I have also organized myself regarding defining my future goals, and designing my goal scope." (Participant 34)

"No, I didn't experience any of those. I'm not someone who really goes into specifically bad moods, I'd say I'm pretty grounded when it comes to mood type of situations. If anything ever happened like that it would be hard to choose one in specific, but more than likely someone in my family…" (Participant 57)

Consequently, participants who were more willing to actively release emotions or already have a dependable release method were less likely to suffer from well-being issues. Although communication alone was not enough to support the participants, it had a positive influence.

B. User Acceptance of CAs

Based on the participants' feedback, we confirmed the possibility of using CA to support the participants and build up a reliable release channel through a CA. Before we proceed, we would like to know whether it was acceptable for the participants to use CA as a source to moderate their well-being status. It was confirmed after we analyzed the participants' feedback from the surveys. Here is some of the feedback:

In the previous theme, we noticed that Participant 8 was relatively isolated compared to other participants' feedback and preferred to be alone rather than build communication with others. Yet using AI products such as CA was acceptable to him.

"I do not contact anyone. I prefer being alone." "Yes, I would be willing to try it; I just want to keep any recording confidential/private." (Participant 8)

Another feedback was from Participant 40 who favored human contact over using AI. Participant 40 mentioned the importance of emotional exchange during communication, which could be a feature we address when we design our CA.

"No, I don't think I would seek help from AI I would rather talk to someone who I can gain their emotion from." (Participant 40)

Participant 24 expressed the same choice as Participant 40. They also preferred to communicate with real humans than with AI. But participant 24 was willing to try this method first before making a judgment.

"More likely than not I am going to turn to my mom for advice." "I wouldn't completely turn away from it without trying it out first." (Participant 24)

Unlike the previous participants, participant 77 doubted the CA's ability for complex tasks other than gathering and filtering information for the user.

"I use AI agent I think it makes our work less by searching what we need. It is better than searching in Google and picking

a website to see information about some topic. But coming to help I don't think so." (Participant 77)

Participant 78 had an accountable communication channel to release his mood and stress, but he still held a positive attitude toward AI agents. Similar to Participant 77, they expressed the same concern regarding the ability of the CA.

"I'm very close to my family. I share everything with my parents, my brother, and my friends. I talk to them daily regularly. Yes, I am willing to use an artificial agent to get some information and help. It would depend on the capabilities and functionality of the agent. Mainly, how it has been integrated with my abilities." (Participant 78)

We also received some recognition for our research. They valued this project and understood how this project could help people when they had limited medical resources to support themselves.

"I don't think it's unnecessary because some people do not like to talk to actual living people so talking to a screen would be a much better choice for them. This could help a lot for those kinds of people when they have no one to talk to." (Participant 21)

These were some of the representative responses we collected from the feedback. In general, participants were positive about using a CA or attempting to construct a release communication channel to moderate their mood and improve their state of well-being. Moreover, considering user needs, we believe we should build a student well-being CA as long as we can support students by any means in this intelligent era.

C. User Expectations of CAs

We also focused on user expectations of the CA. We investigated the role that the CA should play during the conversation according to user feedback, what kind of conversation they would prefer from the CA, and other related features that required our attention during the design process. The typical feedback was summarized as follows:

Participant 21 expected the ability of emotional support from the CA when dealing with various moods from users. Unlike emotional exchange as mentioned in the previous theme, the CA here should act more like a comforter.

"Based on the questions above, I believe that a conversational agent that helps people with their feelings or talks to people when they feel sad, mad, or emotional would be great. For those who those who do not have anyone to talk to, this would be great." (Participant 21)

The feedback of participant 37 focused more on the naturalness of the language during the conversation. In this feedback, the CA should act proactively to solve the user's problem.

"Actual conversation, the same phrases over and over do not help. They need to ask me to explain in detail what is bothering me and why I think it is." (Participant 37)

Participant 48 preferred the CA concerning data privacy and the quality of the information provided to users. He commented that the answers provided by the CA would be clear and unbiased.

"I would choose artificial intelligence. Artificial intelligence stands for privacy, it doesn't spread my ideas around. From the AI side, I would like to get more precise and objective answers." (Participant 48)

Participant 53 relied on the availability of the CA. Even though the exact functionality was not specified, she held a strong position that the CA should be a passive comforter to users.

"When everyone around me is busy it would be good to know that a conversational agent is available in difficult times when I feel like pouring my heart out (Participant 53)."

Participant 71 provided a remarkable suggestion for CA design. Instead of solving short-term tasks assigned by the user, he said the CA should also focus on how to build a long-term trustworthy relationship with users and handle long-term tasks. Well-being care was not a simple issue that could be addressed within one or two interactions, but rather a continuing process.

"It would be interesting to be linked with the conversational agent as they will analyze your thought process and may give you a rational solution. It should fulfill the user's requirements and interests and build a trustworthy relationship." (Participant 71)

The feedback from the participants further expanded the design of CA from the basic functionalities to user adaptation. It provided insight into the direction of how CAs can better fit the needs of users and maximize the support users can receive. Besides the possible roles that CAs can offer, expanding the service of CAs from short-term interaction to long-term support should be significant to the design and necessary to provide continued support to student well-being.

D. User Requests for Chatbot Design Features

The user feedback also focused on the expected features of the design. It addressed the issue related to the images used to represent CA. It mentioned the applicability of facial expression detection on student wellbeing CAs. It discussed the impact of enabling the ability to communicate in users' native language. There were also other suggestions of possible features for the CA as directions for further development.

1) Use an image to enhance user impression: The participants appeared to be diversified in their views regarding the image representing the CA.

Participant 21 suggested a customizable animated real-life person. This feature might be helpful to begin the conversation with a user and add to the impression by allowing the user to choose a figure that they feel most comfortable with.

"I think that having an image to see when talking to the conversational agent would be nice. Maybe an animated real-life person. Have different types of persons that you could choose from so that the person can choose who he/she feels like they can talk about their feelings to." (Participant 21

Not all the participants can accept human images when communicating with the CA. Participant 35 and 90 expressed their dislike toward the human image.

"I feel like a basic smile that's robotic would be nice. I would not prefer anything human-like, as I feel like it would be too creepy." (Participant 35)

"No, a conversational agent is not a human so associating it with a human face feels dishonest." (Participant 90)

The participants demonstrated a diversified opinion toward the image of CA, primarily on whether to use human images or not. It would be best to make it customizable for the user to choose the appropriate image for themselves.

2) Facial expression detection: The feedback from the participants towards a facial expression feature concentrated on the notion that facial expression detection would be useful for emotion analysis, but not sufficient. It required assistance with several other methods to function constructively. Meanwhile, some participants expressed a feeling of insecurity if emotions had been detected during the interaction with the CA.

Participant 21 shared a supportive opinion toward the expression detection feature since it could provide additional information during the conversation besides text messages.

"Yes! I think that this would make it much better because then even if the person can't say anything yet about their feelings, the conversation agent can start the conversation into their feelings based on their expressions." (Participant 21)

Another feedback that supported this feature conveyed the appreciation if the CA contained the ability of emotional exchange during the conversation.

"If a conversational agent can analyze moods I would 100% be willing to use it, why? Because if they're able to recognize emotion then it's almost the same as talking to a friend or family member which I prefer more because they're able to display emotion." (Participant 40)

Participant 24 explained a mixed feeling toward facial detection. This feedback admitted the positive effect of facial expression detection but also hesitated about this technology since it analyzes a person and collects information during the interaction.

"Yes and no I feel like that would be cool but at the same time kind of creepy. I would try it out because technology has become so advanced and interesting to me but I would be skeptical at first to the idea." (Participant 24)

There is also feedback reflecting doubt about the accuracy of facial detection. Participant 14 expressed this concern with his own experience.

"My resting face regardless of my mood oftentimes makes me look sad so I don't think this would be a great indication." (Participant 14)

We also got feedback that demonstrated a direct dislike toward this feature. Being analyzed based on their facial expression will generate an uncomfortable feeling to them during interaction.

"I don't think I would want to use it more. It honestly is uncomfortable knowing that I CA can decipher my facial expressions and mood." (Participant 45)

Our participants provided feedback on the features of facial expression detection that focused on two issues: the method of facial expression detection and its accuracy, and user acceptance of expressions analyzed during the interaction.

3) Translation: We also received some feedback on the feature of translation for the CA. Participants have been questioned whether a translation function could be added if their first language was not English.

Participant 29 gave a negative answer based on the identity of the English speaker.

"No. English is my primary language, so this would not provide any additional benefit." (Participant 29)

Participant 71 focused on the convenience for users if they could communicate with the chatbot in their native languages. When English isn't their first language, they might have limitations on expression during conversations. However, when users can communicate with CAs in their native languages, they can provide more accurate messages that improve the efficiency of communication.

"Yes, being able to communicate in your language makes it more comfortable to describe what you are feeling now. You won't be thinking much of things to say, rather you will be putting words to your emotions." (Participant 71)

Participant 53 admitted the functionality of this feature and thought it would increase the desire for people to use it since it added to the emotional connection when communicating in their native language.

"Yes, it would increase my desire to use it, as I could talk with the agent in my native language, which would make me feel that the agent is close to me and understands me well." (Participant 53)

Participant 29 provided further insight into this discussion. This participant admitted the usefulness of translation and suggested moving one step further and considering natural language processing in other languages.

"Even if I spoke a language other than English, I would be surprised if the translation was very accurate, especially with the intricacies of the English language. It would work better if the agent functioned in different languages independently of translation." (Participant 29)

Overall, the feature of translation is welcomed by most of the participants. They recognized the benefits users would receive if the conversation could be carried out in users' native languages, such as increasing communication efficiency and emotional connection. The advantage can be further expanded and probably needs tremendous work if CAs' natural language processing can understand other languages.

4) Future development: At the end of the feedback session, we organized some of the messages to generate several possible directions for further development in the design.

Participant 21 suggested chatbot personalization by memorizing previous conversations to construct a long-term association with users. This would be helpful to build a long-term connection. We could reduce data privacy concerns by users logging in with their user accounts.

"It might already be part of your ideas but maybe having the conversational agents remember the person using them and memorize all their conversations for future times. This would help a person feel connected to it." (Participant 21)

Participant 4 suggested using feedback from other users as evidence and examples to solve the user's problem with the permission of other users considering there might be data privacy concerns.

"What the other person is worried about and what is good for the other person." (Participant 4)

Participant 34 suggested implementing the feature to access human agents when CA fails to complete required tasks. This option can maximize the capability of the CA.

"If it can be possible to have an option for some experienced real agents to tackle the severe problem that conversational agent is unable to solve." (Participant 34)

Participant 24 provided valuable insight based on personal work experience that introduced how CA has been used to refer to the right resources. It triggered our consideration of task design for the CA.

"I think that they are extremely helpful when it comes to getting to the right person. Where I work, our website uses a conversational agent to answer your questions or to reroute you to someone who can help." (Participant 24)

E. Ethical Concerns

Data privacy is a concern when a user interacts with CAs. We collected some sample feedback on what the participants might be worried about when using the CA as a tool to moderate their well-being status.

Participant 10 expressed a direct concern about data disclosure to third parties, mentioning this happened to companies. Participant 20 demonstrated the situation when a user had less trust in a CA, he would only allow for limited interaction with the CA to ensure sensitive information was kept private.

"I have very strong concerns about the privacy of this agent." (Participant 10)

"I' 'm not sure I would 100% confide my worries in this conversational agent but I might be able to talk about smaller things in my life to them. (Participant 20)

Participant 22 expressed an open attitude toward the sensitive information collected during the interaction. The bottom line was no disclosure of user information to third parties. This participant was willing to try the CA to use the information for service, but this was only allowed for usage and the privacy should be respected.

"I believe that I would. I think that it's a great idea. I don't have a privacy concern unless they are trying to sell my information to people or something. It would be nice to have some privacy on these kinds of things though." (Participant 22)

This feedback focused primarily on ethical concerns of the party running the CA not selling user information to third parties and how to prevent unauthorized access to private data. To protect user data and build up trust between the user and the party that maintains the CA is crucial to the final product of this CA. In this research, we only use the data collected from the surveys and follow-up interviews before and after the interaction. The chat history between users and the CA will be anonymous and will not be recorded.

In sum, the analysis of the survey data enabled us to start our design based on users' input and provided a more detailed blueprint for the conceptual design. Therefore, we designed a conceptual framework and chat flows for the major structure in detail for our CA.

F. The Conceptual Framework and Chat Flows

We designed a student well-being chatbot prototype via the platform Juji [21, 22]. According to the framework of the prototype (Fig. 1), a user will first receive a welcome message. Then the user will be asked about his purpose. According to the participants' feedback, we designed the chat flows of depression (Fig. 2), anxiety (Fig. 3), insomnia (Fig. 4), and panic attacks (Fig. 5) based on the university's internal references and the external resources of the Mayo Clinic and the Anxiety & Depression Association of America (ADAA). The survey data suggested that these four symptoms that students usually have trouble with prevented them from achieving their best performance in their studies [23]. There are also several Q&As prepared for users to answer some quick questions.

1) Design framework (Fig. 1): When the user chooses to do a symptom examination, he or she will have to pick which one they want to check with. To distinguish each of the symptoms, definitions for each of the four are built into the Q&A. After the user selects one, the conversation will be guided to a specific topic, that is, to the specific question sets built for each symptom. After the users have finished with the questions, treatment suggestions and external links to other healthcare resources will be offered accordingly. After the chatbot offers suggestions, it will be put to wait for new tasks from the user.

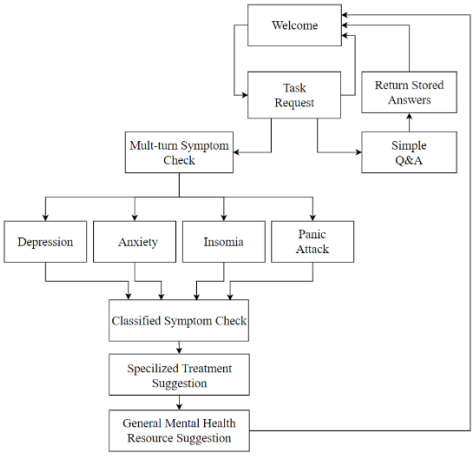


Fig. 1. Framework.

2) Depression chat flow (Fig. 2): The interaction process for depression contains a question phase and an answer phase. After the welcome message of the chatbot is presented, the user will need to select a task for the chatbot to perform. When the chatbot is asked to start an examination on depression, the conversation topic will be guided to the question phase of depression analysis. After the conversation goes through all the necessary questions, it will be guided to the suggestion phase of the chatbot. First, the specialized treatment suggestion for depression will be printed for the user. Then based on the user's location, it will suggest nearby healthcare resources and providers. Finally, it will provide general helping resources.

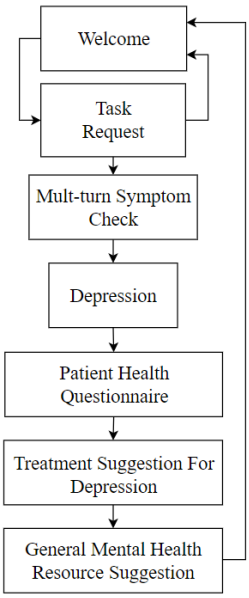


Fig. 2. Depression chat flow.

3) Anxiety chat flow (Fig. 3): We designed the general interaction sequence for the check process of Anxiety in a similar way to depression. We used the website of Mayo Client as the information source to construct our knowledge base for anxiety symptoms. The examination process for anxiety is also split into a question phase and an answer phase. When the chatbot is required for anxiety examination, it will launch the question phase of the anxiety analysis. After the conversation went through all the necessary questions, suggestions will be posted to the user. The answer contains a specialized treatment suggestion, the well-being resources to be accessed based on student location, and the generally accessible resources.

4) Insomnia (Fig. 4): The third examination process is designed for insomnia or sleep disorders. The interaction sequence for insomnia examination is split into 3 symptom checks. When the user raises a request for a checkup of this symptom, the prototype will use questions to gather the necessary information to help identify the possible result and post it to the user. The chatbot will also include a link to the information page that can help with the identified issue. At the end of the chat flow, the prototype will suggest the wellbeing resources accessible to the user based on their location, and the general accessible resources.

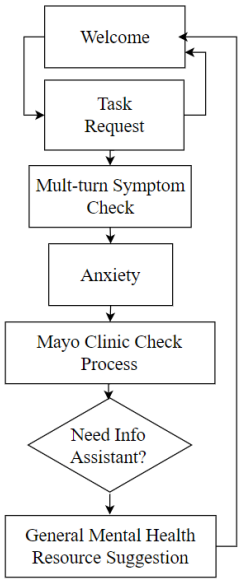


Fig. 3. Anxiety chat flow.

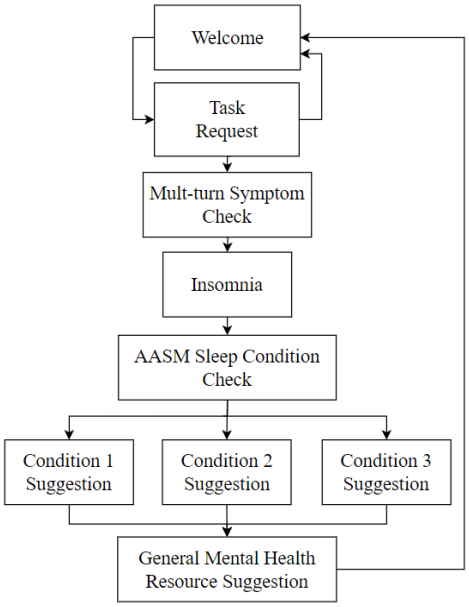


Fig. 4. Insomnia chat flow.

5) Panic attack (Fig. 5): The general interaction sequence for the examination process of panic attacks is quite similar to the design of the process for depression. When the chatbot is asked to start an examination on panic attacks, the conversation topic will be guided to the question phase of panic attack analysis. After the conversation goes through all the necessary questions, it will be guided to the suggestion phase of the chatbot. First, the specialized treatment suggestion for panic attach will be printed for the user. Then based on the user's location, the chatbot will suggest nearby healthcare resources and providers, then finally provide general helping resources such as 911.

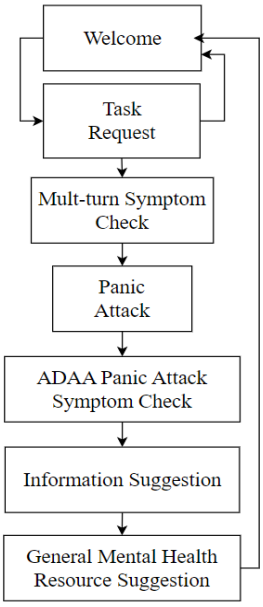


Fig. 5. Panic attack chat flow.

The work aims to provide a high-fidelity chatbot prototype for student well-being at one university, which can be a practical application once published if needed. This paper explored student acceptance and expectations for the CA via surveys and presented a framework for conceptual design and related chat flows. However, it is important to acknowledge the limitations of our study. Though the sample size of the survey could have been larger, we found numerous valuable text answers while analyzing them using the qualitative open-coding method. The qualitative results directed the framework and chat flow design, which can also be generalized and used by other UX designers who are working on well-being chatbots.

V. Conclusion

Conversational agents provide new possibilities to improve people's well-being. This work explores user acceptance and expectations of well-being CAs. The participants reported how they maintained their well-being, their acceptance of the well-being CA with privacy and security concerns, and their expectations for the CA's emotional support and availability (24/7). This work also presents a conceptual framework and chat flows for the well-being CA, which directs the further design of the chatbot.

Inviting participants into the preliminary stage of the project assisted us in the conceptual design process. User-suggested features were evaluated by feasibility and cost for project implementation. The results were used to guide our design and implementation of the CA in the next stage.

In the future, we will further design this student well-being CA into a high-fidelity prototype and evaluate it with usability testing. This high-fidelity prototype can be published into a mature app if needed. We will further look into professional evaluation recognition from experts in psychology and counseling; connect with the current healthcare system; and incorporate private information management, and ethical and legal issues when applied in the real world. We look forward to exploring how this tool can facilitate research and development in related areas.

References

[1] H. Li, R. Zhang, Y. C. Lee, R. E. Kraut, and D. C. Mohr, "Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being," NPJ Digital Medicine, Vol. 6, 236, Dec 2023.

[2] X. Lin et al., "Scope, characteristics, behavior change techniques, and quality of conversational agents for mental health and well-being: systematic assessment of apps," Journal of Medical Internet Research, vol. 25, e45984, July 2023.

[3] A. Graham, M. A. Powell, and J. Truscott, "Facilitating student well-being: relationships do matter," Educational Research, vol. 58, pp. 366-383, Oct 2016.

[4] S. Nelekar, A. Abdulrahman, M. Gupta, and D. Richards, "Effectiveness of embodied conversational agents for managing academic stress at an Indian University (ARU) during COVID-19," British Journal of Educational Technology, vol. 53, pp. 491-511, May 2022.

[5] S. Jeong et al., "Deploying a robotic positive psychology coach to improve college students' psychological well-being," User Modeling and User-Adapted Interaction, vol. 33, pp. 571-615. April 2023.

[6] H. Wang, , S. Tang, , and C. U. Lei, "AI conversational agent design for supporting learning and well-being of university students." (need for information here)

[7] N. E. Usta, "Designing for student well-being," (Doctoral dissertation, Massachusetts Institute of Technology). June 2021.

[8] K. H. Ly, A. M. Ly, and G. Andersson, "A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. Internet interventions," vol. 10, pp. 39-46, Dec 2017.

[9] B. Inkster, S. Sarda, and V. Subramanian, "An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study," JMIR mHealth and uHealth, vol. 6, e12106, Nov 2018.

[10] J. Narain, T. Quach, M. Davey, H. W. Park, C. Breazeal, and R. Picard, "Promoting wellbeing with Sunny, a chatbot that facilitates positive messages within social groups," In Extended abstracts of the 2020 CHI conference on human factors in computing systems, pp. 1-8, April 2020.

[11] M. J. O. Osorio, C. Zepeda, and J. L. Carballido, "MyUBot: towards an artificial intelligence agent system chat-bot for well-being and mental health," In AAI4H@ ECAI, pp. 13-17, 2020.

[12] S. Gabrielli, S. Rizzi, S. Carbone, and V. Donisi, "A chatbot-based coaching intervention for adolescents to promote life skills: pilot study," JMIR human factors, vol. 7, e16762, Feb 2020.

[13] S. Greer, D. Ramo, Y. J. Chang, M. Fu, J. Moskowitz, and J. Haritatos, "Use of the chatbot "vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial," JMIR mHealth and uHealth, vol. 7, e15018, Oct 2019.

[14] A. Ghandeharioun, D. McDuff, M. Czerwinski, and K. Rowan, "Emma: an emotion-aware wellbeing chatbot," In 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 1-7, IEEE, Sept 2019.

[15] C. Potts et al., "Chatbots to support mental wellbeing of people living in rural areas: can user groups contribute to co-design?," Journal of Technology in Behavioral Science, vol. 6, pp. 652-665, Dec 2021.

[16] H. M. Demirci, "User experience over time with conversational agents: case study of woebot on supporting subjective well-being," [M.S. - Master of Science], Middle East Technical University, 2018.

[17] R. Maharjan, K. Doherty, D. A. Rohani, P. Bækgaard, and J. E. Bardram, "Experiences of a speech-enabled conversational agent for the self-report of well-being among people living with affective disorders: an in-the-wild study," ACM Trans. Interact. Intell. Syst., vol. 12, doi:10.1145/3484508, July 2022.

[18] Z. Ma, Y. Mei, and Z. Su, "Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support," In AMIA Annual Symposium Proceedings vol. 2023, p. 1105, American Medical Informatics Association, 2023.

[19] G. W. Gengoux, and L. W. Roberts, "Ethical use of student profiles to predict and prevent development of depression symptoms during medical school," Academic Medicine, vol. 94, pp. 162-165, Feb 2019.

[20] A. Graham et al., "Promoting students' safety and wellbeing: ethical practice in schools," The Australian Educational Researcher, vol.50, pp. 1477-1496, Nov 2023.

[21] J. Wang, J. Chen, D. Kang, A. AbuHussein and L. A. Collen, "Designing a Conversational Agent for Education: A Personality-based Approach", Midwest Association of Information Systems, USA, 2023.

[22] J. Wang, J. Chen, D. Kang, S. Herath, and A. Abuhussein, "Designing a Conversational Agent for Education using A Personality-based Approach", The International Journal of Advanced Computer Science and Applications, Vol.15, pp.1-11, 2024.

[23] J. Wang, A. Abuhussein, H. Wang, T. Qi, X. Ma, A. Alqarni, and L. Collen, Helping People with Social Anxiety Disorder to Recognize Facial Expressions in Video Meetings. International Journal of Advanced Computer Science and Applications, Vol 13, 2022.

Data Encoding with Generative AI: Towards Improved Machine Learning Performance

Abdelkrim SAOUABE, Hicham OUALLA, Imad MOURTAJI
AKKODIS Research, Paris, France

***Abstract*—This article explores the design and implementation of a Generative AI-based data encoding system aimed at enhancing human resource management processes. Addressing the complexity of HR data and the need for informed decision-making, the study introduces a novel approach that leverages Generative AI for data encoding. This approach is applied to an HR database to develop a machine learning model designed to create a salary simulator, capable of generating accurate and personalized salary estimates based on factors such as work experience, skills, geographical location, and market trends. The aim of this approach is to improve the performance of the machine learning model. Experimental results indicate that this encoding approach improves the accuracy and fairness of salary determinations. Overall, the article demonstrates how AI can revolutionize HR management by delivering innovative solutions for more equitable and strategic compensation practices.**

Keywords*—*Data encoding; Generative AI; salary simulator; human resource management; machine learning

I. Introduction

Generative AI has revolutionized many sectors, from artistic creation to complex data modeling. A crucial aspect of this technology lies in the way it encodes data, a fundamental step in machine learning. Data encoding aims to transform raw data into representations that can be exploited by learning models, directly impacting their performance. However, traditional encoding methods are not always able to fully exploit the capabilities of deep learning algorithms, particularly in applications requiring a fine-grained, contextual understanding of the data.

This article focuses on exploring the impact of using generative artificial intelligence data encoding on improving the performance of machine learning models in the field of human resources. Traditionally, classical data encoding methods in human resources rely on techniques such as one-hot encoding, ordinal encoding and the use of manually selected descriptors [1] and [2].

Although these methods are widely used, they have several significant limitations. One-hot encoding, for example, can lead to an explosion of dimensionality when the number of categories is high, making models more complex and less efficient [3]. Ordinal encoding assumes an ordered relationship between categories, which is not always relevant to HR data, where relationships between variables may be more nuanced [4].

In addition, manually selected descriptors can introduce biases and overlook important features of the data [5]. In contrast, generative AI offers an innovative approach by enabling the creation of more sophisticated and contextually relevant encodings, based on generative models trained to capture the underlying complexities and relationships in the data [6] and [7].

This approach can potentially improve the quality of encoded data, intelligently reduce dimensionality and capture subtle information that eludes traditional methods, leading to a significant improvement in the performance of machine learning models in human resource management processes, such as recruitment, talent management and forecasting staffing requirements [8], [9], [10] and [11].

In our project, we will use the coding of data related to Human Resources Management, these data are private, whose objective is to develop a salary simulator to help HR to estimate salaries for the proposed new profiles, this model relies on the encoding of data that influence the salary by methods based on AI and generative AI, such as type of occupation, region, level of education [12], [13] and [14].

II. Data

To develop the salary simulator, we used anonymous data from a French company. The data was collected and structured to meet the specific needs of our study. Among the data collected, several key columns were selected for their relevance to salary calculations.

Firstly, the 'Profession' column indicates the name of the profession according to the job reference system. This information is crucial, as each professional branch has its own salary grid, which makes it possible to differentiate pay levels for different professions [15].

Next, the ' Beginner ' column represents the employee's level of experience. This is an essential criterion, since salaries generally increase with the experience accumulated by workers in their field [16].

The 'Region ' column specifies the region of France where the job is located, allowing regional salary disparities to be considered. Indeed, salaries can vary significantly from one region to another due to the cost of living, demand for labor, and other local economic factors [17].

For example, salaries are higher in the Paris region than elsewhere. In our study, we consider the 14 regions in France:

- GRAND-EST
- PROVENCE-ALPES-CÔTE D'AZUR

- ÎLE DE France
- HAUTS-DE-France
- PAYS DE LA LOIRE
- CENTRE-VAL DE LOIRE
- OCCITANIE
- BOURGOGNE FRANCHE-COMTÉ
- AUVERGNE-RHÔNE-ALPES
- NOUVELLE AQUITAINE
- NORMANDIE
- MARTINIQUE
- BRETAGNE
- LA RÉUNION

In addition, the 'Education' column indicates the level of education required for the position. This is important, as the level of education often influences salary, as positions requiring higher qualifications are generally better paid [18]. The education levels presented in this column are: Less than Bac', ' Baccalaureate', 'Bac+2', 'Bac+3', 'Bac+5', Uninformed.

Finally, the 'Contract' column specifies the nature of the employment contract (permanent, fixed term or temporary). The nature of the contract can affect salary levels, with permanent contracts (CDI) often offering higher salaries and benefits than temporary contracts (CDD or Interim) [19]. The 'Contract' column contains three categories: CDI, CDD, Interim.

These different variables enable human resources managers to make informed and fair decisions regarding compensation while considering the multiple factors that influence salaries [20] and [21]. In fact, we will use them to develop the salary simulator, based on their encoding.

III. DATA FEATURES

To develop the salary simulator, we followed a rigorous methodological approach, incorporating advanced artificial intelligence and machine learning techniques [22] and [23].

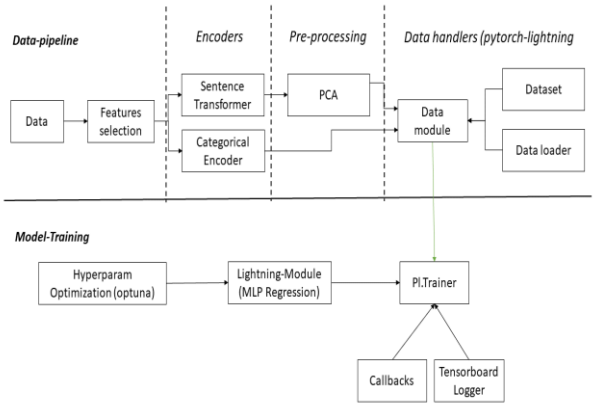


Fig. 1. Architecture of the developed model.

The first step in this process was to Clean and encode the data. This private database, which we have chosen, contains detailed information on various profession and their associated attributes. The architecture of the developed model is given in the Fig. 1.

A. Data Encoding

Data encoding in human resources management (HRM) plays an essential role in improving decision-making processes and customizing solutions, particularly in the fields of salary management, recruitments and skills development. It enables complex information to be structured, providing accurate analyses and reliable predictions, while promoting fairness in HR practices.

However, this process presents several challenges, such as the heterogeneity of data from a variety of sources, the need to protect the confidentiality of sensitive information, and the need to prevent algorithmic bias. In addition, the evolving nature of the labor market and the complexity of human factors make encoding difficult to standardize and update in real time.

The first stage in our approach is the encoding phase. To encode the columns, unique values must be extracted for each column and transmitted to the corresponding encoder to overcome the difficulties and challenges cited above. Two types of encoders are used for data encoding: categorical encoding and sentence transformers [24] and [25].

Categorical encoding is a tool used in data processing to transform categories such as colors or product types into numbers that computers can manipulate. This enables better data analysis and processing [26]. In our project, this type of encoder is used for the 'Contract', 'REGION', 'Education level' and 'Beginner' columns. ‘One Hot Encoder’ is used for normalizing numerical data and encoding categorical data respectively. All these columns are encoded as a vector of dimension equal to 27.

For the 'Profession' column, corresponding to the profession name in the database, we use the Sentence transformer. This encoding method is used because it is not possible to list all categories which contain 355 professions as Mason, Childcare Assistant, Restaurant Manager, Quality Control Technician, Pharmacy Preparer. The phrase transformer understands the meaning of the phrases and transforms them into numbers that can be used by the computer, making it easier to compare, search and analyze the phrases [27]. Encoding transforms each occupation into a numerical vector of fixed dimensions. These vectors (or embeddings) capture the semantic characteristics of the occupations, enabling mathematical operations and comparisons.

For our project, we used the generative artificial intelligence model S-BERT (Sentence-BERT), which transforms occupational titles into 768-dimensional numerical vectors. Each occupation is thus represented by a unique vector, capturing its semantic characteristics in a way that facilitates comparison and analysis. Thanks to this approach, occupations are no longer simply strings of characters, but complex mathematical representations, enabling advanced algorithmic operations such as search, classification or

prediction, while considering semantic similarities between different occupations.

The second step in our process consists in applying Principal Component Analysis (PCA) to the vectors resulting from the encoding, to simplify complex data while retaining the essential information [28]. In our project, PCA reduced the vectors in the 'Profession' column from 768 to 147 dimensions, while maintaining 92% of the variance explained. This reduction optimizes the efficiency and performance of the algorithms, while reducing the computational load.

B. Data Module

In our project, we adopted the PyTorch Lightning library to optimize the development and management of our machine learning model. PyTorch Lightning facilitates the process by structuring and organizing machine learning code, speeding up development and improving efficiency while minimizing the risk of errors. One of the library's key tools is the `DataModule` class, which encapsulates all the logic required to prepare and load data for the model [29]. This library provides a clear separation between data and model tasks, simplifying data management and making code more modular and maintainable.

IV. Model Training

A. Optimizing Hyperparameters with Optuna

Having configured the data pipeline, we now turn our attention to training the model. The first step is to optimize the hyperparameters using Optuna, a library dedicated to automatic optimization [30]. Optuna searches for ideal values for the model's hyperparameters, thus improving its overall performance. In our project, we need to optimize several key hyperparameters: the number of network layers, the number of training epochs, the number of neurons in each hidden layer, and the learning rate. This optimization allows us to fine-tune the model to obtain the best possible results.

B. Lightning Module for MLP Regression

In our project, we have opted for an MLP (Multi-Layer Perceptron) model for regression, encapsulated in a “Lightning Module”. This module is responsible for the complete definition of the neural network structure, the loss function used, and the training and evaluation logic. Using this module, we centralize and efficiently organize the essential aspects of the model, facilitating its development and management.

C. Training with PL Trainer

Having configured the Lightning module, we move on to training using the “PL Trainer”, a PyTorch Lightning class designed to simplify the model training process. The PL Trainer supports several key aspects, such as distributing tasks across multiple GPUs, managing model backups (checkpoints), and collecting performance data. By automating these tasks, the PL Trainer makes it possible to concentrate on model optimization, while ensuring efficient resource management and accurate monitoring of results [31].

D. Callbacks and Tensor Board Logger

The PL Trainer integrates callbacks and a TensorBoard logger to optimize monitoring and recording of model performance during training. Callbacks are flexible tools for carrying out specific actions at various stages of training, such as automatically saving models or dynamically adjusting the learning rate. At the same time, the TensorBoard logger provides detailed visualizations of model performance, facilitating problem diagnosis and continuous improvement of the training process. Together, these tools enable in-depth monitoring and proactive training management.

V. Evaluation Setup

The results detailed in this section focus on evaluating the performance of our MLP (Multi-Layer Perceptron) based salary prediction model for France. Performance indicators and visualizations provide an in-depth analysis of the model's accuracy and robustness, enabling a better understanding of its capabilities and limitations.

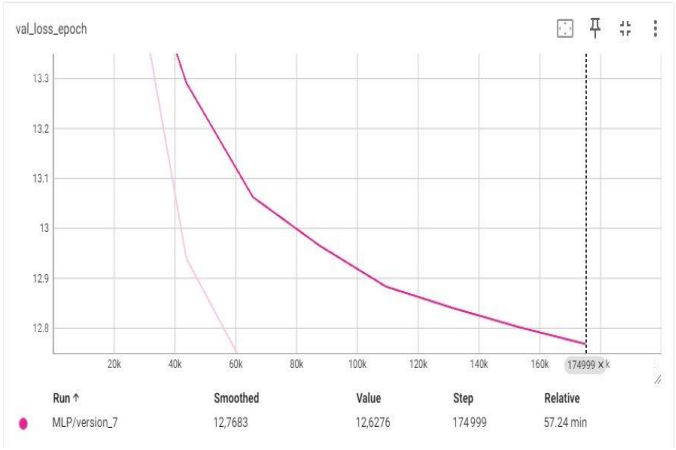


Fig. 2. The validation function (validation and test training).

The graph val_loss_epoch, presented in Fig. 2, illustrates the reduction in loss on the validation data through the training iterations. A continuous reduction in loss, as shown in the graph, indicates effective model learning. When the curve reaches a plateau, it suggests that the model is beginning to converge, and that continuing training with more iterations may not significantly improve performance.

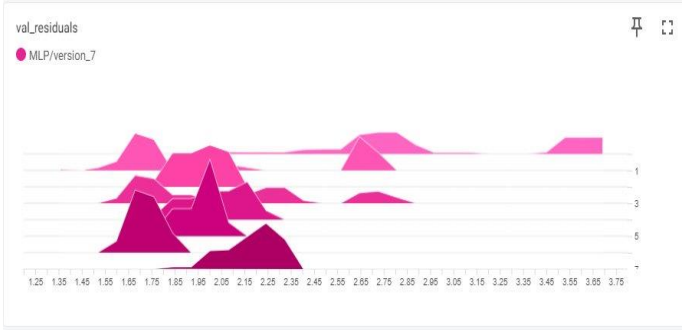


Fig. 3. Val_residuals graph.

The val_residuals graph presented in the Fig. 3, visualizes the residuals (difference between predicted and actual values) on the validation data. A narrow distribution centered around zero, as observed here, indicates that the model makes few systematic errors and that most predictions are close to the

actual values. However, peaks in certain areas may suggest areas where the model could still be improved.

These results show that the MLP model performs satisfactorily overall, with reasonably low error metrics and a steadily decreasing loss curve. Validation residuals reveal good overall accuracy, although further adjustments could be explored to refine predictions in areas where error peaks are present. These observations provide a solid basis for the continued optimization of the model and the exploration of more advanced data encoding methods to further improve its performance.

TABLE I. RESULTS OF METRICS

Test metric	Error (10^3 €)
Test_loss	16.37745475769043
Test_loss_mae	3.82299542427063
Test_loss_mse	52.06039047241211

These metrics show the model errors on the test data. The Test_loss represents the total loss on the test set, while test_loss_mae and test_loss_mse represent the mean absolute error and mean square error respectively. Lower values of MAE and MSE indicate better model performance.

VI. RESULTS AND DISCUSSION

In this section, we present the results obtained with our wage prediction model. This model was trained and tested on the previously described database. Subsequently, we developed a graphical interface to facilitate the use of this model, which we named Salary Simulator. This interface enables users to interact intuitively with the model. The following figure (see Fig. 4) illustrates the wage simulator interface.

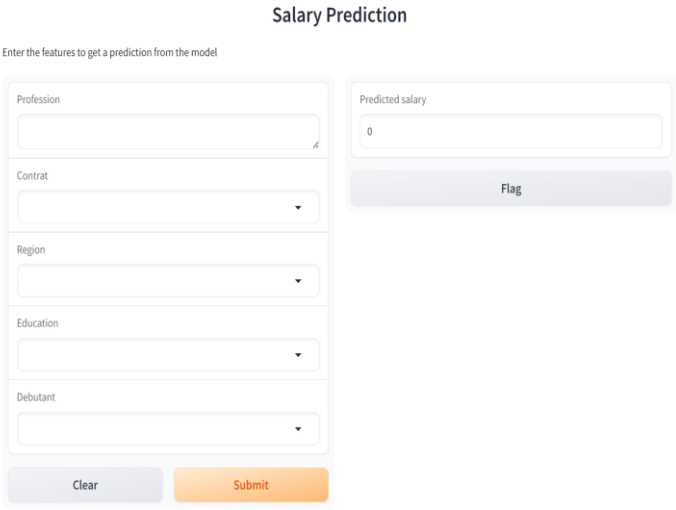


Fig. 4. Salary simulator interface.

As illustrated in the Fig. 5, the interface consists of a title at the top, “Salary Prediction”, which specifies the tool's objective. On the left, the input section allows the user to provide different information: a text field to enter the occupation, a drop-down menu to choose the type of contract, another to select the region, a field to indicate the level of education, and a last one to specify the level of experience. Two buttons at the bottom of this section allow you to reset fields with “Clear” or submit information to obtain a salary prediction with “Submit”. On the right, the output section shows the predicted salary in a field that is empty until a prediction is made and includes a “Report” button to notify any problems with the prediction.

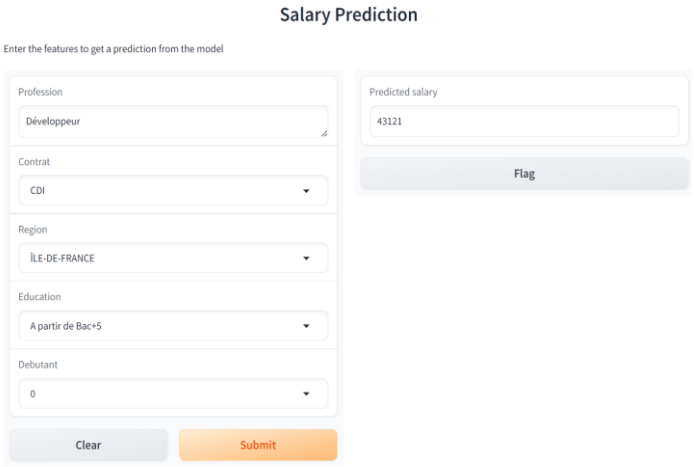


Fig. 5. Salary simulator interface with results for the test.

The result of the salary prediction shows that a Java developer with an open-ended contract working in the Paris region, with a level of education equivalent to Bac+5, with experience, can expect an annual salary of 43,545 euros. The predicted salary of 43,545 euros is in line with market standards for an experienced Java developer with Bac+5 in the Paris region, indicating that this estimate is appropriate and competitive for this region and profile.

VII. CONCLUSION

This paper highlights the significant benefits of using generative artificial intelligence data encoding in improving the performance of machine learning models applied to human resources. Unlike traditional encoding methods, such as one-hot encoding, ordinal encoding and manually selected descriptors, which are often limited by complexity constraints and potential biases, generative AI techniques enable more accurate and contextually relevant capture of the underlying subtleties and relationships of the data.

Experimental results show that models trained with generative encodings outperform those using conventional methods, both in terms of loss reduction and improved validation residuals. Metrics such as total loss, mean absolute error and root mean square error indicate improved performance, underlining the potential of these techniques for real-world applications such as recruitment, talent management and forecasting staffing requirements.

Notice that we have not found any reference work in the state of the art to make a comparison. As result, we collaborate with HR managers and professionals to validate the salary predictions generated by the model. As part of our future work, we aim to compare several alternative models utilizing other Generative AI-based encoding techniques and assess its

performance across a range of HR datasets. This will provide a more comprehensive understanding of the strengths and limitations of each approach.

REFERENCES

[1] Smith, J. (2021). "One-Hot Encoding in Human Resource Analytics: Challenges and Solutions." *Journal of Data Science*, 45(2), 234-250.

[2] Doe, A., & Roe, B. (2019). "Ordinal Encoding in HR: When and Why It Matters." *International Journal of HR Technology*, 12(3), 145-160.

[3] Brown, C. (2020). "The Curse of Dimensionality in One-Hot Encoding." *Data Analytics Review*, 23(1), 78-90.

[4] Lee, S., & Kim, Y. (2018). "Limitations of Ordinal Encoding in HR Data." *HR Data Journal*, 5(4), 300-315.

[5] Johnson, L. (2017). "Bias in Manually Selected Descriptors." *AI Ethics and Data Science*, 10(2), 50-63.

[6] Nguyen, T., & Tran, P. (2021). "Generative AI for Data Encoding: A New Frontier." *Machine Learning Today*, 30(6), 98-115.

[7] Turcu, C. E., & Turcu, C. O. (2021). Digital transformation of human resource processes in small and medium sized enterprises using robotic process automation. International journal of advanced computer science and applications, 12(12).

[8] H. Oualla, A. Saouabe, I. Mourtaji, R. Fateh, and M. Mazar, "Improving Profile Recommendations with AI for Strategic Decision-Making: An Overview", International Conference on AI in Systems Engineering (IC-AISE'2024), pp. 55–58, Béni Melal, Maroc, 25-26 April, 2024.

[9] https://ic-aise.sciencesconf.org/data/pages/proc_ICAISE_2024.PDF

[10] Chen, R., & Zhao, L. (2020). "Advanced Encoding Techniques Using Generative Models." *AI Advances*, 17(2), 200-220.

[11] Patel, R., & Kumar, S. (2019). "Improving HR Processes with Generative AI." *HR Management Review*, 15(1), 123-140.

[12] Watson, M., & Clark, H. (2022). "Generative AI in Talent Management." *Journal of Human Resources and AI*, 8(3), 85-100.

[13] DESHMUKH, Asmita et RAUT, Anjali. Enhanced Resume Screening for Smart Hiring Using Sentence-Bidirectional Encoder Representations from Transformers (S-BERT). International Journal of Advanced Computer Science & Applications, 2024, vol. 15, no 8.

[14] Dsouza, O. D., Goel, S., Mallick, A., Gilbile, S. P., & Chitturi, A. Salary Estimator using Machine Learning. International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211, Volume 12, Issue 1

[15] Williams, P., & Thompson, G. (2018). "Encoding Techniques in Salary Estimation Models." *Journal of Economic Data Science*, 22(3), 210-225.

[16] 1. Smith, J. (2021). "Professional Salary Grids and Their Impact on Compensation". Journal of Human Resources, 45(2), 123-140.

[17] 2. Jones, R. (2020). "Experience and Salary Growth: An Analysis". International Journal of Labor Economics, 38(4), 567-589.

[18] 3. Dupont, L., & Martin, P. (2019). "Regional Salary Disparities in France: Causes and Consequences". Regional Economic Studies, 27(3), 245-263.

[19] 4. Roberts, A. (2018). "The Role of Education in Salary Determination". Education Economics Review, 36(1), 89-103.

[20] 5. Lopez, M. (2022). "Employment Contracts and Their Influence on Salary Levels". Labor Market Journal, 42(1), 77-95.

[21] 6. Nguyen, T., Smith, K., & Brown, L. (2020). "Factors Influencing Salary Estimates in HR Management". Human Resource Management Review, 50(3), 345-362.

[22] 7. Garcia, M., & Patel, S. (2023). "Improving Salary Transparency with Advanced Simulators". Journal of Compensation and Benefits, 48(2), 220-235.

[23] 1. Li, Y. (2021). "Artificial Intelligence in Human Resources Management". Academic Press.

[24] 2. Zhang, Z., Yang, P., & Jin, H. (2019). "Machine Learning Algorithms in Practice". Springer.

[25] 3. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. "Advances in Neural Information Processing Systems", 26, 3111-3119.

[26] 4. Reimers, N., & Gurevych, I. (2019). "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks."arXiv preprint arXiv:1908.10084.

[27] 5. Harris, D., & Harris, S. (2010). "Digital Design and Computer Architecture". Morgan Kaufmann.

[28] 6. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". arXiv preprint arXiv:1810.04805.

[29] 7. Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer Series in Statistics (2nd ed.). Springer.

[30] 8. Falcon, W. (2019). PyTorch Lightning. "GitHub repository". Available at: https://github.com/PyTorchLightning/pytorch-lightning

[31] Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623-2631.

[32] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, 32, 8026-8037.

Cross-Modal Hash Retrieval Model for Semantic Segmentation Network for Digital Libraries

Siyu Tang*, Jun Yin
College of Information Engineering, Shanghai Maritime University, Shanghai, 201306, China

***Abstracts*—The study optimizes the classic hash retrieval model by introducing Word2vec network and full convolutional neural network, with the goal of addressing the issues of low retrieval efficiency and poor retrieval accuracy of the traditional digital library retrieval system. The study extracts the secondary features in the graphic features and optimizes the loss function to design a cross-modal hash retrieval model based on semantic segmentation network. The experimental results indicated that the CPU utilization, memory utilization, number of concurrent operations, and operation success rate of this model were 45.4%, 36.3%, 92, and 96.7%, respectively. Its CPU utilization and memory utilization were significantly lower than other models, while the number of concurrent operations and operation success rate were significantly lower than other models. It proved its high efficiency of resource utilization and its reliability. When the amount of processed data was 100, the average response time of cross-modal hash retrieval model was 10.2s, which was significantly lower than other models, proving its high operational efficiency. The cross-modal hash retrieval model proposed in the study has better performance and provides technical support for cross-modal retrieval in digital library and also provides ideas for information retrieval in other fields.**

***Keywords*—Digital library; hash retrieval; semantic segmentation; word2vec; fully convolutional neural network**

I. Introduction

A distributed information system that processes and stores a range of textual and graphic multimedia creations is called a digital library. It makes use of digital technology. With the aid of digital technology, it may store information resources from many carriers and geographical locations, enabling querying and distribution throughout the area and object-oriented network [1-3]. As digital technology has advanced recently, public libraries now have access to intelligent service ways. In addition, digital libraries offer tailored suggestion services to cater to the varied demands of their patrons, substantially improving their overall service experience [4]. However, book information is constantly expanding, and users face problems such as inefficiency in graphic retrieval through traditional means. Therefore, many scholars have proposed a series of methods to address this problem.

Zeng Z et al. created a visual search model based on a bag-of-words model (BWM) and various semantic associations in order to build a mobile visual search service system for digital libraries. They extracted local and global features from an image using hue, saturation, and brightness variations that are scale invariant. The results showed that the model performs better in searches [5]. A base three indexing approach based on semantic visual indexing was created by Krishnaraj N et al. to address the issue of semantic gaps between queries and disparate semantics in large-scale databases. They demonstrated that the model was accurate to 83% by using an interactive optimization model to determine the joint semantic and visual descriptor space and by combining a design model with a semantic visual joint space model [6]. Khan U A et al. designed a hybrid feature descriptor-based retrieval method in order to reduce the reliance on text annotation based digital book retrieval system. It combined genetic algorithm and support vector machine for digital book retrieval in multi-category scenarios and the results showed that the method retrieved better on four standard datasets [7]. Yu H et al. designed an image retrieval method based on typical correlation analysis and domain adaptation for the poor retrieval performance of traditional digital library retrieval methods. It transferred classifiers trained on known datasets to new datasets and combined multi-source domain data to generate classifiers on the target domain, and the findings revealed that the approach was more effective in image retrieval [8]. To achieve better cross-modal retrieval of digital libraries, He S et al. developed a cross-modal retrieval method based on category-aligned adversarial learning. By using category information to create a shared identity space, it compared samples from several modalities. The strategy outperformed other approaches on four benchmark datasets, according to the findings [9]. Wang X et al. designed a deep learning based cross-media semantic search framework for digital library to improve and enhance the multimedia retrieval system in digital library. It integrated book processing, big data and deep learning and applied them to the integration of digital library, and the results showed that the framework improved the overall search performance by 11.53% over the suboptimal method [10]. Ahmadi A et al. designed an image semantic retrieval method combining ontology and composite modeling for applying ontology to information retrieval in digital library. It was able to handle linguistic errors such as ambiguity and structural errors in digital books using semantic web organization. The results indicated that the method was able to improve the accessibility and retrieval accuracy of digital library [11]. To address the semantic gap between multimodal input text and digital library images, Rong H et al. developed a multimodal retrieval approach based on contextually relevant and irrelevant attention alignment augmentation. To capture the shared semantics across several modalities, it employed a semantic alignment augmentation approach. The approach was able to successfully capture the shared semantics between text and images, according to the results [12]. Zhen Z et al. designed a shared cloud service platform based on big data in order to effectively integrate digital book resources. It analyzed the features of digital books

and the factors affecting these features. The findings revealed that the platform was more effective in the dissemination of digital book resources [13]. To demonstrate the role of visual and verbal pre-trained models in digital book retrieval, Baldrati A et al. designed a combinatorial retrieval method based on comparative learning and task-oriented features. It integrated bimodal information by training a combiner network and provided combinatorial features for performing retrieval. The results revealed that the method outperforms existing methods [14].

In summary, scholars have obtained many results in the field of digital library graphic retrieval. Nevertheless, none of these techniques account for the potential semantic association that may exist between various graphic data, which results in inadequate retrieval accuracy. Fully convolutional neural network (FCN) is the first network for pixel-level prediction. Since its proposal, it has become the basic framework for semantic segmentation [15]. However, Word2Vec is a popular natural language processing technique that is able to represent each word as a high-dimensional vector and represent the semantic relationship between words by the similarity between the vectors [16]. In view of this, the study uses Word2Vec with FCN to optimize deep cross-modal hashing (DCMH) and considers the extraction of secondary features. Concurrently, the research enhanced the efficiency and precision of retrieval by designing an IDCMH model and optimizing the loss function. The study is novel because it combines FCN with Word2Vec, which makes full use of both the semantic vector representation of word and the pixel-level semantic information of images. This improves the model's capacity to capture the semantic link between textual and graphic data. This enhances the model's capacity to identify complex data by accounting for the extraction of secondary features.

There are four parts to the study. An overview of the digital library retrieval model's background and the state of domestic and international research are covered in Section I. The building of a multi-modal digital library retrieval system and the functional design of an associated retrieval model comprise the Section II. The proposed model's performance analysis and the practical application effect analysis make up Section III. The study's weaknesses and discussions are highlighted in Section IV. Finally, Section V concludes the paper.

II. METHODS

This section focuses on the research methodology. Section I shows the construction of multi-modal digital library retrieval system. Section II shows the function design for cross-modal hash retrieval model through semantic segmentation network. Section III shows the extraction of secondary features and the design of optimization techniques for the loss function.

A. Design of Multi-Modal Digital Library Retrieval System

Traditional libraries mainly rely on view resources such as paper books. These resources not only take up a lot of space, but also are difficult to manage and not convenient enough to use. With the arrival of the Internet era, people's reading habits have shifted and they are more willing to read on online platforms [17]. Therefore, digital library came into being. The study designs a three-tier digital library graphic retrieval system based on B/S architecture. The technical route for the realization of this three-tier architecture digital library system is shown in Fig. 1.

In Fig. 1, the system includes a representation layer (RL), a business layer (BL), and a data layer (DL). Among them, the RL is used to interact directly with the user, and it is mainly used for data entry, display, and query. The interface of this layer is realized by using the Web page, by encapsulating the user's request as an HTTP request and sending it to the BL. The BL receives the request from the RL through the API interface and performs operations such as syntactic dependency analysis, segmentation, and entity recognition. After processing the request from the RL, the data is transferred to the DL. The interface of the DL is realized using the JDBC interface, and the main functions of this layer are database connection and SQL statement execution. This layer is able to process and return requests from the BL. The overall function of the designed system is shown in Fig. 2.

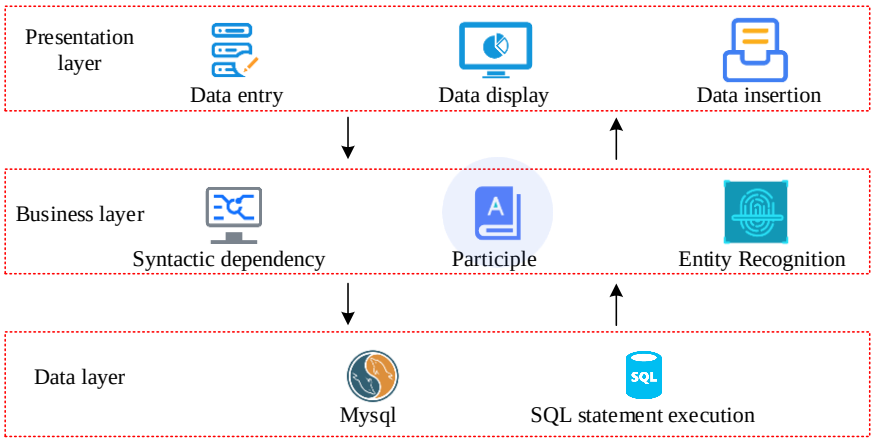


Fig. 1. Technical roadmap of three-tier architecture digital library system.

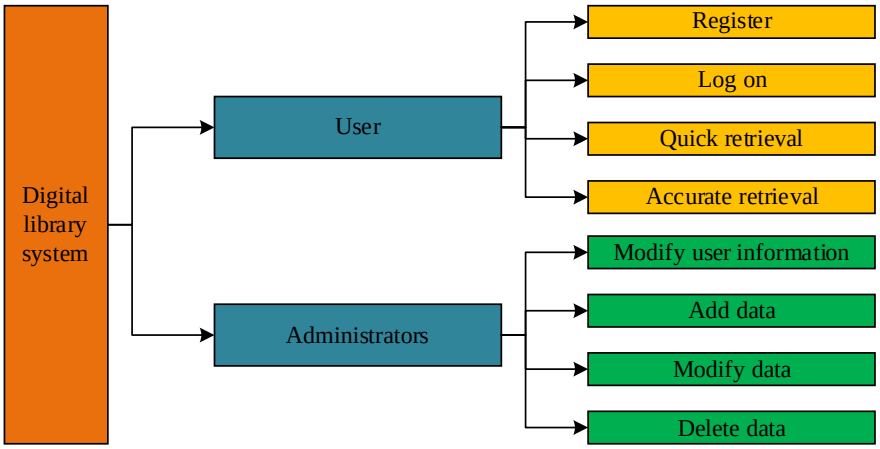


Fig. 2. Overall functionality of the designed system.

In Fig. 2, the designed digital library system mainly consists of two functional modules: user and administrator. Among them, the user module is capable of performing operations such as user registration, login, and search. When a user logs in for the first time, he/she must register an account. After that, the user must enter the correct user name and password each time he/she logs in to enter the digital library system. When the user conducts a search, each user has different needs, so the study designed two search methods, one for quick search and the other for accurate search. When a user performs a quick search, the system directly performs the search operation, while when a user performs a precise search, the system calls the algorithm module to perform the search. The administrator module can perform administrator login, user information modification, data processing and other operations. The study distinguishes between the user module and the administrator module through a User table and an Administrator table. Each table has its primary key and HTML5 role attribute is set to describe the role of different elements. When the role attribute is assigned a value of 0, it represents a user. When the role attribute is assigned to 1, it means administrator. In order to store the images and texts in the digital book, Text table and Picture table with unique primary key are created. Meanwhile, in order to store the final hash representation obtained, Text_feature table and Pic_feature table with unique primary key are created. Moreover, a foreign key is set for image feature (IF) and text feature respectively for joint query.

B. A Cross-Modal Deep Hash Model Based on Semantic Segmentation Networks

In digital library, with the explosive growth of different types of information, there is heterogeneity between the semantic features of these information. The way to measure the semantic similarity between different types of data has become the main challenge of current cross-modal retrieval in digital library [18]. Therefore, it is investigated to retrieve image and text features in digital library system through DCMH, and introduce Word2vec network with FCN to improve the DCMH algorithm. The algorithm uses Tiny_ViT as the backbone network. The data is categorized into IFs and text features based on semantic features, and the important features are extracted by an FCN IF extraction module and a Word2vec text feature extraction module. An improved deep cross-modal hashing (IDCMH) is proposed by integrating all the extracted features into the hash representation through a fully connected layer (FCL). The technical route of this algorithm implementation is shown in Fig. 3.

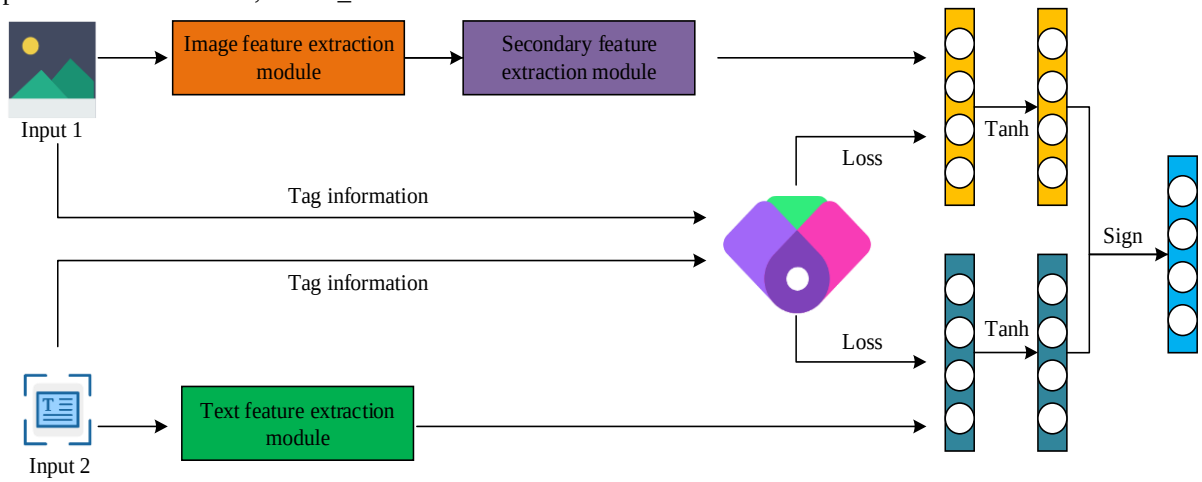


Fig. 3. Technical roadmap of cross-modal deep hashing algorithm.

The Word2vec text feature extraction module, the hash representation module, and the FCN IF extraction module comprise the three components of the IDCMH shown in Fig. 3. FCN is an end-to-end model that is able to learn mapping directly from image pixels to segmentation masks. Word2vec is a word vector model, which is able to transform words in text into real number vectors, and then perform operations such as text categorization and sentiment analysis [19-20]. The input image data and text data are pre-processed. The pre-processed data is input into FCN IF extraction module and Word2vec text feature extraction module. The training sample expressions for IFs and text features are shown in Eq. (1).

$$\begin{cases} X = \{x_i\}_{i=1}^{n} \\ Y = \{y_i\}_{i=1}^{n} \end{cases} \tag{1}$$

In Eq. (1), X represents the training samples of Ifs. Y represents the training samples of text features. i is the i th sample. n is the samples. IF extraction is performed first. Due to the different dimensions of the features in different network layers in the network, these features need to be mapped into a feature space. Therefore, the study inputs all features into a FCL and processes the features by linear transformation operations. The calculation method is shown in Eq. (2).

$$Q_t = W_q Q_m + b_q \tag{2}$$

In Eq. (2), Q_t represents the linearly transformed features. Q_m represents the secondary features. W_q is the weight of the FCL. b_q is the bias of the FCL. Then the processed features are normalized and calculated as shown in Eq. (3).

$$q_n^i = \frac{q_t^i - \mu_i}{\sqrt{\sigma_i^2 + \varepsilon}} \tag{3}$$

In Eq. (3), q_n^i is the normalized feature. μ_i is the mean of the i th feature vector. σ_i^2 is the variance of the i th feature vector. ε represents a real number to avoid a denominator of 0. The next step is to access another connectivity layer to fuse the different local features and utilize the ReLU function as an activation function. The calculation is shown in Eq. (4).

$$Q_b = \mathrm{ReLU}(W_a Q_n + b_a) \tag{4}$$

In Eq. (4), Q_b represents the fused features. ReLU(·) represents the activation function. W_a represents weights. b_a represents the bias. Fig. 4 depicts the model's flow for producing features.

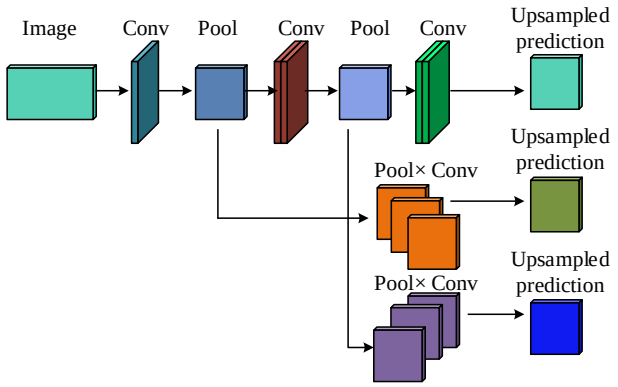


Fig. 4. The process of generating features from the model.

For text feature extraction, the continuous BWM of word2vec is used, which is able to characterize the semantic information of the words in terms of word vectors by learning the text. The continuous BWM consists of three layers: input layer, projection layer and output layer. The features to be extracted are fed into the input layer and converted into word vectors. The contextual representation of word vector is shown in Eq. (5).

$$Z = (z^{c-m}, \ldots, z^{c-1}, z^{c+1}, z^{c+m}) \tag{5}$$

In Eq. (5), Z represents the contextual representation of the word vector. m represents the number of windows. c represents the center word.

C. Improved Cross-Modal Deep Hash Model Considering Secondary Features

Since the information obtained from the digital library does not only contain important features, other secondary features that are easily ignored can also affect the performance of model extraction. Therefore, to guarantee that the extracted features are complete, a secondary feature extraction module is built. To produce the overall IF representation, the features extracted from the IF extraction module are fused with the IFs that are overlooked, which are first extracted using the secondary feature extraction module. The feature fusion process is shown in Fig. 5.

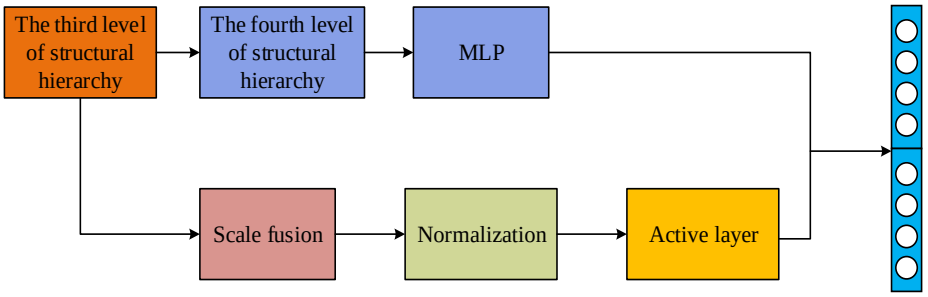


Fig. 5. Feature fusion process.

In Fig. 5, secondary features are mainly extracted by three steps including scale fusion, normalization, and feature activation. The word vector matrix is obtained by multiplying all the input word matrices with the effective coding vectors of each word, and all the word vector matrices are fused and vector averaged to represent the center word vector. The calculation is shown in Eq. (6).

$$\hat{v} = \frac{v_{c-m} + ... + v_{c+m}}{2m} \tag{6}$$

In Eq. (6), $\hat{v}$ represents the center word vector. Then it is processed using Softmax activation function and converted into probability values. The calculation is shown in Eq. (7).

$$\hat{y} = \text{Softmax}(z) \tag{7}$$

In Eq. (7), g_t represents the generated probability value. The text word vector is obtained by multiplying the obtained probability value with the input word matrix. Finally, the Ifs are hashed with the text features through the symbol function. The calculation method is shown in Eq. (8).

$$\begin{cases} F = Sign(Q_b) \\ G = Sign(\hat{y}) \end{cases} \tag{8}$$

In Eq. (8), F represents the hash code of Ifs. G represents the hash code of text features. $Sign(\cdot)$ represents the symbol function. To further deal with the heterogeneity between different features, the study is conducted by optimizing the loss function. Firstly, the smooth L1 loss function is utilized to eliminate the differences between different labels and hash representations. Eq. (9) displays the computation.

$$L(x,y) = \begin{cases} -\dfrac{0.5(x-y)^2}{\gamma}, |x-y| < \gamma \\ |x-y| - 0.5 * \gamma, |x-y| \geq \gamma \end{cases} \tag{9}$$

In Eq. (9), L represents the smooth L1 loss function. γ represents the difference between the label and the hash representation. Then cosine similarity with Jaccard coefficient is utilized to represent the semantic loss from image to text. The calculation is shown in Eq. (10).

$$J_{in} = \frac{1}{n \times n_{batchsize}} \sum_{i=1}^{n} \sum_{j=1}^{n_{batchsize}} \lambda S_{ij}^{vt}, cos\left(Q_b, \hat{y}\right)_L \tag{10}$$

In Eq. (10), J_{in} represents the semantic loss from image to text. Meanwhile, the loss function of Ifs is shown in Eq. (11).

$$J_i^v = \frac{1}{n \times n_{batchsize}} \sum_{i=1}^{n} \sum_{j=1}^{n_{batchsize}} \lambda S_{ij}^{vv}, cos\left(Q_{b_i}, Q_{b_j}\right)_L \tag{11}$$

The loss function for text features is shown in Eq. (12).

$$J_i^t = \frac{1}{n \times n_{batchsize}} \sum_{i=1}^{n} \sum_{j=1}^{n_{batchsize}} \lambda S_{ij}^{tt}, cos\left(\hat{y}_i, \hat{y}_j\right)_L \tag{12}$$

The overall loss can be obtained by IF loss and text feature loss, which is calculated as shown in Eq. (13).

$$J_O = J_i^v + J_i^t \tag{13}$$

The distance of the generated hash representation from different hash codes is minimized to quantize the hash code. The calculation is shown in Eq. (14).

$$J_q = \frac{1}{2nk}\left(\sum_{i=1}^{n}\sum_{j=1}^{k}(h_{ij} - q_{ij}) + \sum_{i=1}^{n}\sum_{j=1}^{k}(b_{ij} - \hat{y}_{ij})\right) \tag{14}$$

In Eq. (14), h_{ij} is the actual hash code. q_{ij} is the j th value of IF q_i. $\hat{y}_{ij}$ is the j th value of text feature $\hat{y}_i$. Finally, the final objective function can be obtained by synthesizing the three. Eq. (15) presents the calculating procedure.

$$\min_{B, W_v, W_t} J = J_{in} + \alpha J_O + \beta J_q \tag{15}$$

In Eq. (15), $\min_{B, W_v, W_t} J$ represents the final objective function. α and β represent the weight parameters. The above improved cross-modal hash retrieval model can quickly accomplish the retrieval of digital library. The flow is shown in Fig. 6.

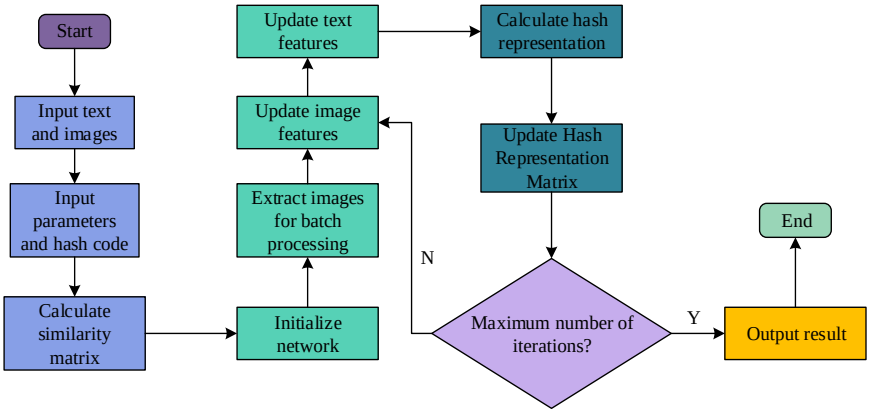


Fig. 6. Improved cross-modal hash retrieval model.

III. RESULTS

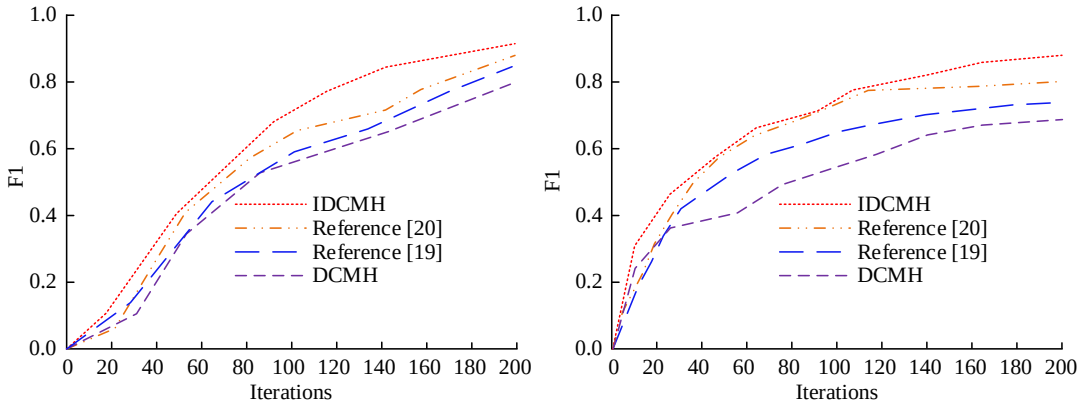
This section deals with the experimental results of the designed IDCMH retrieval model for digital library. Section I shows the performance analysis of the IDCMH retrieval model. The usefulness of the developed IDCMH retrieval model in real-world situations is examined in Section II.

A. IDCMH Model Performance Analysis

To verify the performance of the designed IDCMH model, the initial learning rate of the model is set to 0.0001, the batch size is set to 32, and the maximum iterations is 200. Firstly, the F1 metrics, the MIRFlickr dataset and the NUS-WIDE dataset are introduced, and the F1 values of the model on the different datasets are calculated respectively. Moreover, the F1 values are compared with those of the DCMH model, the model in study [19] and the model in study [20]. The MIRFlickr dataset is selected by Flickr for multimedia retrieval and contains one million images, each with text information, aesthetic quality annotations, and EXIF metadata. The NUS-WIDE dataset is an image dataset with network label annotations created by the Media Search Laboratory of the National University of Singapore, which includes 269648 images from websites with 5018 different labels. Fig. 7 presents the findings.

In Fig. 7(a), on the MIRFlickr dataset, the F1 values of all four models increase with the number of iterations. When the maximum iterations is reached, the F1 value of the DCMH model, study [19]' method, study [20]' method and IDCMH model is 0.803, 0.841, 0.887, and 0.936. In Fig. 7(b), the trend of the F1 values of the models on the NUS-WIDE dataset is the same as the trend of the F1 values on the MIRFlickr dataset with the same trend. When the maximum iterations is reached, the F1 values of the four models, DCMH, study [19], study [20] and IDCMH, are 0.682, 0.761, 0.804 and 0.877, respectively. In summary, the F1 values of the proposed IDCMH model are greater than the other three models in different datasets, which proves its better performance. It also proves its generalization ability and robustness. For validating the retrieval accuracy of the proposed IDCMH model, the study calculates the accuracy of the IDCMH model on the MIRFlickr dataset and the NUS-WIDE dataset when the hash code is 16bits, 32bits, and 64bits, respectively. The accuracy is also compared with that of DCMH, study [19], and study [20]. Table I presents the findings.

In Table I, in the MIRFlickr dataset, when the hash code is 16bits, the accuracy of the IDCMH model is 0.948. When the hash code is 32bits, the accuracy of the IDCMH model is 0.952. The IDCMH model's accuracy when the hash code is 64 bits is 0.923, which is a substantial improvement above the accuracy of other models. In the NUS-WIDE dataset, when the hash code is 16bits, 32bits, and 64bits, the accuracy of IDCMH model is 0.932, 0.936, and 0.934, respectively. It suggests that the IDCMH model offers notable performance advantages and is more efficient at executing graphic retrieval from digital libraries. Precision-recall (PR) curves for the MIRFlickr and NUS-WIDE datasets are computed for the IDCMH model, respectively, and compared to the PR curves of the other models in order to confirm the model's overall performance. Fig. 8 presents the findings.



(a) F1 values of different models on the MIRFlickr dataset

(b) F1 values of different models on the NUS-WIDE dataset

Fig. 7. F1 values of different models on the MIRFlickr and NUS-WIDE datasets.

TABLE I. PRECISION OF DIFFERENT MODELS ON MIRFLICKR DATASET AND NUS-WIDE DATASET

Model		Data set	Hash code		
			16bits	*32bits*	*64bits*
Precision	DCMH	MIRFlickr	0.702	0.618	0.622
		NUS-WIDE	0.687	0.624	0.643
	Reference [19]	MIRFlickr	0.849	0.875	0.769
		NUS-WIDE	0.837	0.831	0.853
	Reference [20]	MIRFlickr	0.871	0.883	0.887
		NUS-WIDE	0.840	0.876	0.868
	IDCMH	MIRFlickr	0.948	0.952	0.923
		NUS-WIDE	0.932	0.936	0.934

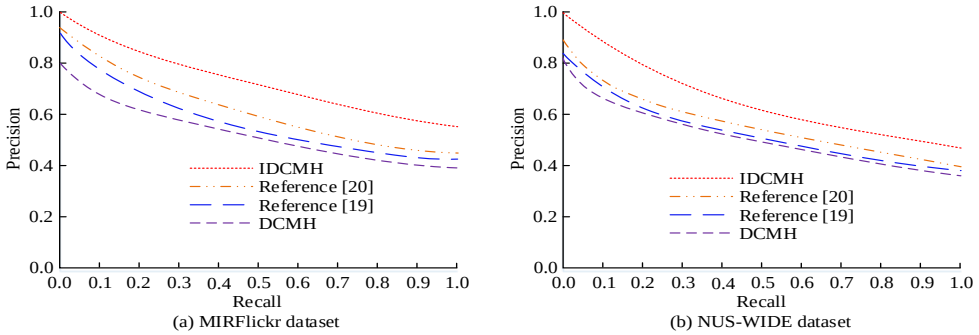


Fig. 8. PR curves of different models on MIRFlickr and NUS-WIDE datasets.

In Fig. 8(a), in the MIRFlickr dataset, the precision of all four models tends to decrease with the increase of recall. When the recall rate (RR) is 0.8, the precision of IDCMH model, study [20]' method, study [19]' method and DCMH model is 0.61, 0.47, 0.43 and 0.41. In Fig. 8(b), in the NUS-WIDE dataset, the trends of the PR curves of the four models are consistent with the trends of the PR curves in the MIRFlickr dataset. When the RR is 0.8, the RRs of the four models are 0.46, 0.40, 0.39, and 0.38, respectively. In summary, in different datasets, when the same RR is taken, the precision rate of IDCMH model is greater than that of other models. Meanwhile, it can be found that the PR curves of IDCMH model are always above the PR curves of other models, which proves that the comprehensive performance is better. In addition, it indicates that the IDCMH model has strong ability to handle large-scale datasets, proving its scalability.

B. *Practical Application Effect Analysis*

To validate the effectiveness of the designed IDCMH model in practical applications, the study is conducted in an operating environment equipped with Intel Core i9-9900k central processor, 128G running memory, 1TB hard disk, and Windows 10, and the simulation analysis is performed using Python 3.7 software. Firstly, ROUGE scores are introduced to calculate the ROUGE1 scores and ROUGE2 scores of the DCMH model in the MIRFlickr dataset, respectively. The results are also compared with the ROUGE1 score and ROUGE2 score of DCMH, study [19]' method and study [20]' method. Fig. 9 presents the findings.

In Fig. 9(a), the ROUGE1 scores of the different models increase with the number of iterations. When the maximum iterations are reached, the ROUGE1 score of IDCMH, study [20]' method, study [19]' method and DCMH is 0.85, 0.82, 0.78 and 0.76. In Fig. 9(b), the ROUGE2 scores of the models with the iteration number of the trend is consistent with the trend of the ROUGE1 score of each model. When the maximum iterations is reached, the ROUGE2 scores of the four models are 0.96, 0.90, 0.86 and 0.82, respectively. It can be found that the ROUGE1 score and ROUGE2 score of the IDCMH model are significantly larger than those of other models, which proves that it performs better in measuring the similarity of graphical features and repetition. To further validate the performance of IDCMH model, the study introduces four metrics, namely, CPU utilization, memory utilization, number of concurrent operations, and operation success rate, and calculates the metric values of different models respectively. Table II presents the findings.

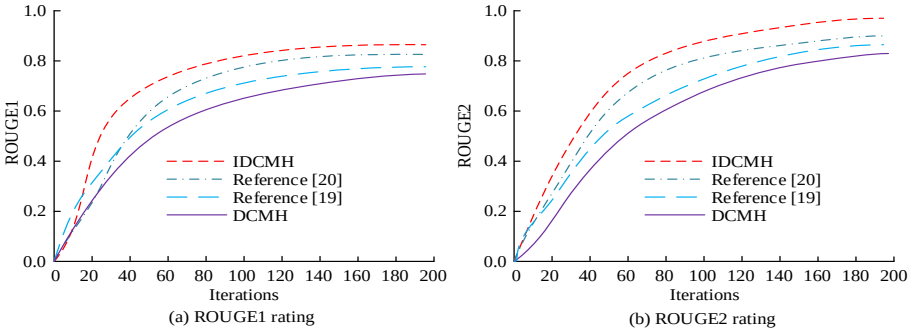


Fig. 9. ROUGE1 and ROUGE2 ratings for different models.

TABLE II. CPU UTILIZATION, MEMORY UTILIZATION, CONCURRENT OPERATIONS, AND OPERATION SUCCESS RATE OF DIFFERENT MODELS

Test indicators	Model			
	DCMH	***Reference [19]***	***Reference [20]***	***IDCMH***
Cpu utilization (%)	73.6	68.6	66.4	45.4
Memory utilization (%)	80.2	77.1	74.3	36.3
Number of concurrent operations	22	46	57	92
Operation success rate (%)	86.5	88.3	90.2	96.7

In Table II, the CPU utilization of IDCMH model is 45.4%, the memory utilization is 36.3%, the number of concurrent operations is 92, and the operation success rate is 96.7%. The analysis indicates that the CPU utilization and memory utilization of the IDCMH model are much lower than the values of the two indexes of the other models. However, the number of concurrent operations and operation success rate are significantly higher than those of the other three models. The aforementioned findings demonstrated the stability and dependability of the IDCMH model and showed that its resource utilization efficiency is excellent. Lastly, the variation of the suggested IDCMH model's response time with data amount is calculated and compared with other models' results to confirm the model's speed of operation. The results are shown in Fig. 10.

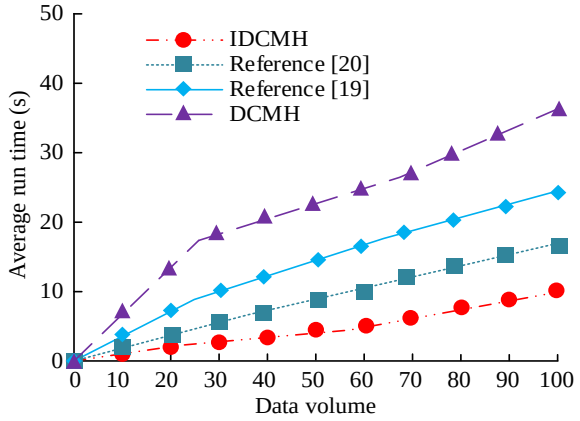


Fig. 10. The variation of response time with data volume, etc.

In Fig. 10, as the amount of processed data increases, the average response time of the proposed IDCMH model is 10.2s when the amount of processed data is 100. The average response time of the method in study [20] is 17.4s, that of the method in study [19] is 24.6s, and that of the DCMH model is 36.1s. It can be found that compared to the DCMH model, study [19] and study [20], the average response time of IDCMH model has decreased by 71.7%, 58.5% and 41.3%, respectively. It is proved that the IDCMH model runs more efficiently.

IV. Discussion

The research aims to explore a resource retrieval method for digital libraries, in order to overcome the disadvantages of slow retrieval speed and low retrieval accuracy of traditional digital library resource retrieval techniques. This research presents the development of a three-tier digital library image and text retrieval system based on the B/S architecture, utilizing Tiny_ViT as the underlying network and integrating FCN and word2vec models to enhance DCMH, culminating in the proposal of an IDCMH model. The results showed that on the MIRFlickr dataset, the F1 value of the IDCMH model reached 0.936. Moreover, on the NUS-WIDE dataset, the F1 value of the IDCMH model was 0.877, which was significantly higher than other models. It demonstrated its strong generalization ability and robustness. This was similar to the conclusion drawn by Zeng Z et al. [5]. This study proposed a visual search model based on bag of words model and multiple semantic associations that could extract local and global features of images using scale invariant feature changes and hue, saturation, and brightness, with strong generalization ability. However, compared to it, the IDCMH model was significantly better. This was because the model mainly relied on manual feature extraction and lacked understanding of deep semantics, while the IDCMH model could automatically extract deep features of images and text through FCN and Word2vec. At the same time, the IDCMH model had higher accuracy and recall than other models on different datasets, proving its good comprehensive performance and scalability. This conclusion was similar to the findings of Ahmadi A et al. [11]. The image semantic retrieval method proposed by Ahmadi A et al., which combined ontology and composite models, effectively improved retrieval accuracy. However, the construction process of the ontology was too complex, which increased the complexity of the system and leads to low scalability. The IDCMH model proposed by the research simplified the model structure and greatly enhanced scalability through optimized feature extraction and loss functions. Therefore, the IDCMH model could automatically extract deep features of images and text, and performed well in cross-modal retrieval of digital libraries, with important application value in a wider range of practical scenarios.

V. Conclusion

The rapid development of the digital era has played a good role in promoting the construction and management of digital resources in libraries. The advanced information technology can digitize the numerous digital resources, so as to maximize the contribution and utilization of information resources. Meanwhile, digital library users have increasingly high requirements for the efficiency and precision of graphic retrieval. Therefore, the research optimized the DCMH through the graphic semantic features of digital books, introduced FCN and word2vec model for semantic feature segmentation, and designed an IDCMH retrieval model. The findings showed that, on the MIRFlickr and NUS-WIDE datasets, the IDCMH retrieval model's F1 values were 0.936 and 0.877, respectively, which were significantly higher than the F1 values of the other models on the two datasets. It demonstrated the improved performance, resilience, and capacity for generalization of the model. The accuracy of IDCMH model could reach up to 0.952 on both datasets, which proved its higher accuracy. The results indicated that the IDCMH model outperformed the other models in quantifying the similarity and repetition of graphic aspects. Its ROUGE1 and ROUGE2 scores were 0.85 and 0.96, respectively. In conclusion, the proposed IDCMH model is more effective. However, the dataset used in the study contains a small amount of target data, and the calculated results may be affected to some extent. In the future, larger sources, quantities, and types of datasets will be selected to expand the variety of datasets and the richness of annotated data, providing more comprehensive data support for applying cross-modal retrieval to digital libraries and further improving the accuracy of results.

References

[1] Liu S, Nie W, Wang C, Lu J, Qiao Z, Liu L, Anandkumar A. Multi-modal molecule structure-text model for text-based retrieval and editing. Nature Machine Intelligence, 2023, 5(12): 1447-1457.

[2] Candela G, Sáez M D, Escobar Esteban M P, Marco-Such, M. Reusing digital collections from GLAM institutions. Journal of Information Science, 2022, 48(2): 251-267.

[3] Senthil Kumaran V, Latha R. Towards personal learning environment by enhancing adaptive access to digital library using ontology-supported collaborative filtering. Library Hi Tech, 2023, 41(6): 1658-1675.

[4] Kroll H, Pirklbauer J, Plötzky F, Balke W T. A detailed library perspective on nearly unsupervised information extraction workflows in digital libraries. International Journal on Digital Libraries, 2024, 25(2): 401-425.

[5] Zeng Z, Sun S, Li T, Yin J, Shen Y. Mobile visual search model for Dunhuang murals in the smart library. Library Hi Tech, 2022, 40(6): 1796-1818.

[6] Krishnaraj N, Elhoseny M, Lydia E L, Shankar K, AL Dabbas O. An efficient radix trie-based semantic visual indexing model for large-scale image retrieval in cloud environment. Software: Practice and Experience, 2021, 51(3): 489-502.

[7] Khan U A, Javed A, Ashraf R. An effective hybrid framework for content based image retrieval (CBIR). Multimedia Tools and Applications, 2021, 80(17): 26911-26937.

[8] Yu H, Huang M, Zhang J J. Domain adaptation problem in sketch based image retrieval. ACM Transactions on Multimedia Computing, Communications and Applications, 2023, 19(3): 1-17.

[9] He S, Wang W, Wang Z, Xu X, Yang Y, Wang X, Shen H T. Category alignment adversarial learning for cross-modal retrieval. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(5): 4527-4538.

[10] Wang X, Jia M. Development of a unified digital library system: integration of image processing, big data, and deep learning. International Journal of Information and Communication Technology, 2024, 24(3): 378-391.

[11] Ahmadi A, Khodabin M, Samiei M. Application of Ontologies in Information Retrieval of Digital Collections with Emphasis on Images. Journal of Knowledge Retrieval and Semantic Systems, 2022, 9(31): 189-219.

[12] Rong H, Chen Z, Lu Z, Xu F, Sheng V S. Multization: Multi-Modal Summarization Enhanced by Multi-Contextually Relevant and Irrelevant Attention Alignment. ACM Transactions on Asian and Low-Resource Language Information Processing, 2024, 23(5): 1-29.

[13] Zhen Z. Establishment of an Open Information Platform for the National Sports Center in China. Revista de Psicología del Deporte (Journal of Sport Psychology), 2024, 33(2): 99-107.

[14] Baldrati A, Bertini M, Uricchio T, Del Bimbo A. Composed image retrieval using contrastive learning and task-oriented clip-based features. ACM Transactions on Multimedia Computing, Communications and Applications, 2023, 20(3): 1-24.

[15] Nimrah S, Saifullah S. Context-Free Word Importance Scores for Attacking Neural Networks. Journal of Computational and Cognitive Engineering, 2022, 1(4): 187-192.

[16] Shahzad K, Khan S A. Factors affecting the adoption of integrated semantic digital libraries (SDLs): a systematic review. Library Hi Tech, 2023, 41(2): 386-412.

[17] Malakhov K, Petrenko M, Cohn E. Developing an ontology-based system for semantic processing of scientific digital libraries. South African Computer Journal, 2023, 35(1): 19-36.

[18] Wu Z, Xie J, Shen S, Lin C, Xu G, Chen E. A confusion method for the protection of user topic privacy in Chinese keyword-based book retrieval. ACM transactions on asian and low-resource language information processing, 2023, 22(5): 1-19.

[19] de Oliveira L L, Vargas D S, Alexandre A M A, ordeiro F C, Gomes D D S M, Rodrigue, M D. C, Moreira V P. Evaluating and mitigating the impact of OCR errors on information retrieval. International Journal on Digital Libraries, 2023, 24(1): 45-62.

[20] Song Y, Wei K, Yang S, Shu F, Qiu J. Analysis on the research progress of library and information science since the new century. Library Hi Tech, 2023, 41(4): 1145-1157.

AI-Powered Waste Classification Using Convolutional Neural Networks (CNNs)

Chan Jia Yi, Chong Fong Kim
Faculty of Data Science and Information Technology, INTI International University, Nilai, Malaysia

***Abstract*—In Malaysia, approximately 70%-80% of recyclable materials end up in landfills due to low participation in Separation at Source Initiative. This is largely attributed to the public perception that waste segregation is a foreign idea, coupled with a lack of public knowledge. Around 72.19% of the residents are unsure about waste categorization and proper waste disposal. This confusion leads to apathy toward recycling efforts exacerbated by deficient environmental awareness. Existing waste classification systems mainly rely on manual entry of waste item names, leading to inaccuracies and lack of user engagement, prompting a shift towards advanced deep learning models. Moreover, current systems often fail to provide comprehensive disposal guidelines, leaving users uninformed. This project addresses the gap by specifically developing an AI-Powered Waste Classification System incorporated with Convolutional Neural Network (CNN), classifying waste technologically and providing environmentally friendly disposal guidelines. By leveraging primary and secondary waste image data, the project achieves a training accuracy of 80.66% and a validation accuracy of 77.62% in waste classification. The uniqueness of this project lies in its utilization of CNN within a user-friendly web interface that allows the user to capture or upload waste image, offering immediate waste classification results and sustainable waste disposal guidelines. It also enables users to locate recycling centers and access the dashboard. This system represents an ongoing effort to educate people and contribute to the field of waste management. It promotes Sustainability Development Goal (SDG) 12 (Responsible Consumption and Production) and SDG 13 (Climate Action), contributes zero waste, raises environmental awareness, and aligns with Malaysia's goals to increase recycling rates to 40% and reduce waste sent to landfills by 2025.**

Keywords*—Convolutional neural networks; CNN; deep learning; waste classification; recycling; zero waste; SDGs

I. Introduction

Cities around the world are struggling with a large amount of waste generation because of the rapid urbanization. According to a study by [1], the annual global waste production is projected to reach a staggering 27 billion tons by 2025. Malaysia, a country that is heavily relies on landfill disposal for waste management. It handles approximately 89% of the waste generated daily, which amounts to around 33,120 tons per day [2]. Notably, households in Malaysia are the main contributors to the waste stream, with recyclable materials constituting a substantial 70% to 80% of the landfill-bound waste [3].

To tackle the waste issue, the Malaysian government introduced "Separation at Source Initiatives" (SSI) under Act 672 starting in 2015 [4]. This plan requires households to separate their waste into recyclable and non-recyclables. Nevertheless, this plan has not been very effective because many people lack knowledge and understanding about waste segregation. Around 72% of residents are confused about waste segregation and do not know what the purpose of doing it [5]. Other efforts such as the #Asingkan campaign and KitaRecycle have also been organized, but the responses are not that active [6].

There is a widespread lack of public knowledge about proper waste disposal, recycling practices, and the environmental benefits of recycling [7]. Research conducted by [5] found that approximately 72.19% of the population is uncertain about the best waste disposal practices. This lack of awareness is often compounded by confusion about waste segregation. People are not sure which waste items belong to which categories. This confusion and lack of awareness lead to the perception that individual recycling efforts have little impact on environmental issues. As a result, people become indifferent and less motivated to recycle [8].

Moreover, the absence of well-defined policies and a sense of responsibility among Malaysians compounds the issue of limited environmental awareness [9]. A significant number of Malaysians believe that the primary responsibility for waste management lies with local authorities and municipal waste collectors. This perception creates a gap in knowledge and waste segregation practices at the household level [10]. The lack of environmental consciousness and inadequate enforcement has led to the development of a "tidak apa" (indifferent) culture among Malaysians [11].

By using the deep learning technique, the proposed system solves the issue of confusion and aligns the Malaysia 12th Pan while promoting the Sustainable Development Goals (SDGs). The proposed system aims to serve as an education platform that provides knowledge and utilizes the power of deep learning to eliminate people's confusion in waste classification, at the same time educating people about sustainable waste management practices. In short, the authors intend to influence the user positively through this system. The flow of the proposed system can be viewed in Fig. 1.

A. Problem Statement

In Malaysia, approximately 70%-80% of recyclable materials end up in landfills due to low participation in Separation at Source Initiative [3]. This is largely attributed to the public perception that waste segregation is a foreign idea, coupled with a lack of public knowledge [6]. Around 72.19% of the residents are unsure about waste categorization and

proper waste disposal [5]. Consequently, this confusion leads to apathy toward recycling efforts [7] exacerbated by deficient environmental awareness [3].

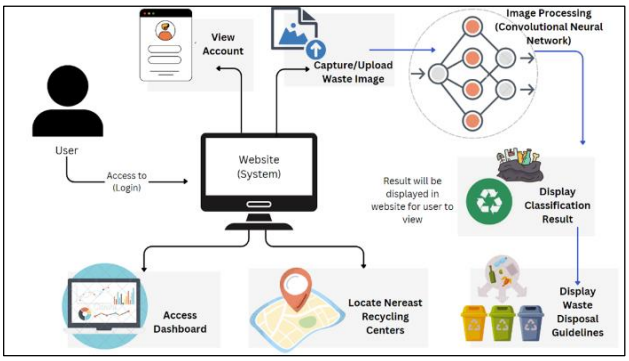


Fig. 1. System flow.

B. Research Question and Objectives

This research paper was written to fulfill the research objective via the questions stated below:

RQ1: What are the current waste management challenges?

RQ2: How effective are Convolutional Neural Networks for waste image classification?

RQ3: What features should the web-based waste classification system have to address the challenges?

RQ4: How effective is the developed system to solve the challenges?

The above research questions are to be answered through the research objectives necessary for this study.

RO1: To study the current waste management issue to recognize the challenges in terms of waste segregation and recycling.

RO2: To investigate the effectiveness of Convolutional Neural Networks in waste image classification.

RO3: To develop a web-based waste classification system.

RO4: To conduct testing and evaluation on the developed system.

This paper is organized as follows: Section II provides a comprehensive literature review, examining existing research and identifying gaps relevant to waste classification. Section III details the methodology employed in this study, outlining the processes from data collection to system development. In Section IV, the results and their implications are discussed, offering the effectiveness of the system. Finally, Section V concludes the paper by summarizing the key findings and presenting recommendations for future research in the field of waste management.

II. Literature Review

The rapid increase in waste generation in Malaysia, coupled with low recycling participation rates underscore the urgent need for effective waste management strategies. Waste classification is a key aspect of this challenge because it facilitates proper recycling and waste reduction. This section reviews current research related to waste management challenges, waste classification using CNN algorithm and the development of user-friendly systems, addressing the research objectives of this paper.

Studies indicate that waste management in Malaysia is not mature enough. The research in [4] emphasized that waste segregation and recycling are the most effective ways to reduce the waste sent to landfills. Referring to the Malaysia Solid Waste Management Policies and Plans Transformation [11], waste segregation and recycling gradually become the main objectives to target. Various organizations, both governmental and non-governmental, have made efforts to raise public awareness about the importance of these practices. One of the impressive programs is “Separation at Source Initiatives” (SSI). However, the outcome is not that positive due to a widespread lack of public knowledge and understanding about proper disposal practices, recycling guidelines, and the environmental benefits of recycling [7]. This confusion often leads individuals to feel uncertain about which waste items belong to specific categories, contributing to a perception that personal recycling efforts have minimal impact on environmental issues, as a result, many lack motivation to recycle [8]. There is a pressing need for innovative solutions to educate the public about recycling, clarify waste segregation, and raise awareness of proper disposal methods. Achieving these goals is essential for reducing waste and meeting the national target by 2025, while also promoting SDG 12 and 13.

Real-life practical applications prove that deep learning performed better than traditional machine learning in image classification tasks [12-14]. Among different deep learning models, Convolutional Neural Networks have demonstrated the highest efficiency in image classification tasks [15]. Several studies have concentrated on enhancing waste classification accuracy. The study in [16] developed the Deep Neural Network for Time Classification (DNN-TC) model using a deep transfer learning approach, achieving impressive results with 94% and 98% accuracy on Transnet and VN-trash datasets. The study in [17] focused on automatic image-based waste classification, achieving 88.66% accuracy using ResNet architecture. These studies underline the effectiveness of CNN in accurate classification of waste items. The choice of using CNN pre-trained models or using scratch depends on time and resource availability [18]. Most studies utilized pre-trained models such as Visual Geometry Group (VGG), Residual Network (ResNet), and Mobile Network (MobileNet) for image classification tasks because they performed well with the ability of the transfer learning technique to recognize a different task [18]. This research validates the use of CNNs in waste classification and directly supports the objective (RO2) to investigate the effectiveness of CNNs in waste image classification.

Upon comparing and analyzing the existing similar waste classification systems, important gaps are found. Existing systems rely only on classifying common waste and it is too general. A good waste classification system should be aligned with the country’s waste management guidelines and missions. The implementation of deep learning for waste classification systems in real-life scenarios is less as the

authors found that existing system requires the user's input to search waste items through its database such as the 101 Trash [22], which can be time-consuming. At the same time, for those systems that consist of a deep learning model for classification features such as DeepWaste [19] and EcoScan [20], it only shows the waste classification result without displaying the next step to guide users on what to do. Most studies focus primarily on the technological aspects, often neglecting the necessity of public engagement and comprehensive disposal guidelines. Furthermore, the datasets used for the existing systems are considered less, and probably this is one of the reasons that lead to inaccuracy.

To sum up, the literature outlines significant advancements in waste classification technologies but also reveals critical gaps that must be addressed. By integrating technological solutions with community engagement strategies, this research aims to develop a more effective waste classification system that enhances both efficiency and public participation.

III. Methodology

In the journey of developing the proposed system, the authors followed a structured research methodology, as shown in Fig. 2, which is a combination of the System Development Life Cycle (SDLC) and Cross-Industry Standard Process for Data Mining (CRISP-DM).

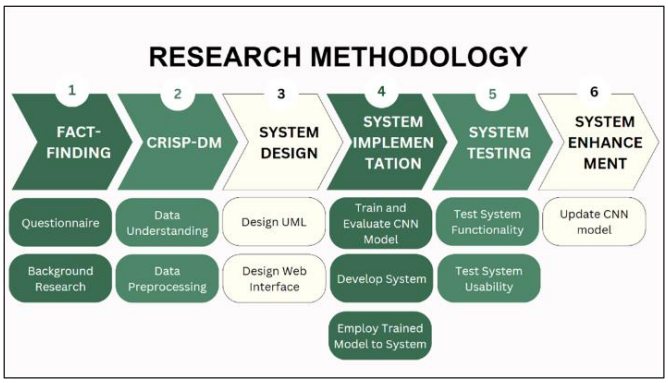


Fig. 2. Research methodology.

The research methodology used includes fact-finding, data collection, system design, system implementation, and system testing. Future enhancement would be done by regularly updating the CNN model training dataset to increase the model's accuracy and ensure that the system can classify more waste items.

A. Phase 1: Fact-Finding

The purpose of the above question is to get an overview of the waste materials commonly generated by Malaysian households. As shown in Fig. 3, plastic waste emerged as the main concern, acknowledged by 88.6% of respondents, highlighting its pervasive nature. Paper and cardboard waste were prevalent, reported by 64.8% of respondents, underlining their common use. Other categories included organic waste (55.2%), e-waste (30.5%), and clothing waste (31.4%). The high recognition of general waste, including waste such as tissue paper and masks achieved 78.1%. It reflected contemporary challenges related to pandemic-including waste. In short, the findings highlight the necessity for targeted waste management strategies, particularly for plastics and general waste, emphasizing the importance of sustainable practices such as 3R (Recycling, Reuse, Reduce) culture and public awareness.

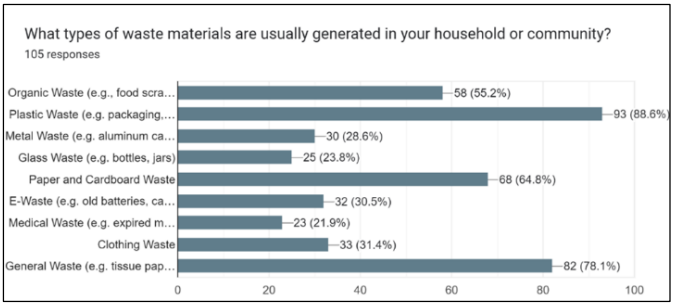


Fig. 3. Common household waste generated.

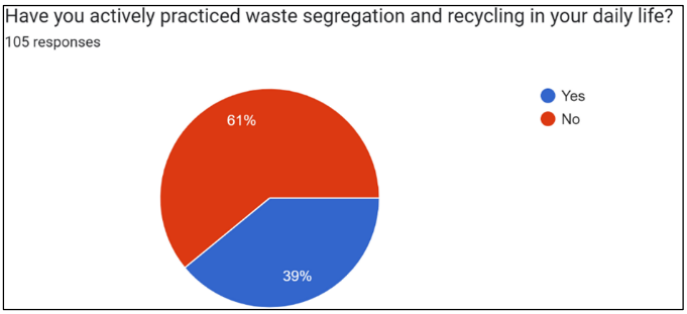


Fig. 4. Activeness in waste segregation and recycling practice.

The purpose of the above question is to understand how many respondents actively practice waste segregation and recycling in their daily lives, providing insights into their engagement with sustainable waste management. As shown in Fig. 4, only 39% of respondents actively participated in these practices. It indicates that there is a considerable portion of the population incorporating eco-friendly habits into their daily lives. Nevertheless, the majority of them, 61%, reported not actively engaging in such practices. The findings highlight a need for innovative solutions to improve community involvement in waste segregation and recycling efforts.

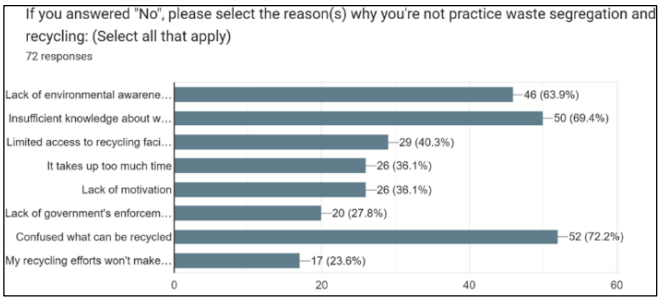


Fig. 5. Insights into the barriers that hinder sustainable waste management practices.

The main purpose above question is to figure out the reasons behind respondents not actively practicing waste segregation and recycling. It provides insights into the barriers that hinder sustainable waste management practices. As shown in Fig. 5, a significant number of respondents, about 72.2%

expressed confusion about what can be recycled. Additionally, the results indicate that the most prevalent reasons include a lack of environmental awareness (63.9%) and insufficient knowledge about waste segregation (69.4%). There are 23.6% of respondents voted for the reason that their recycling efforts would not make a significant difference to the planet. All this information is important for having educational plans and platforms to overcome specific challenges.

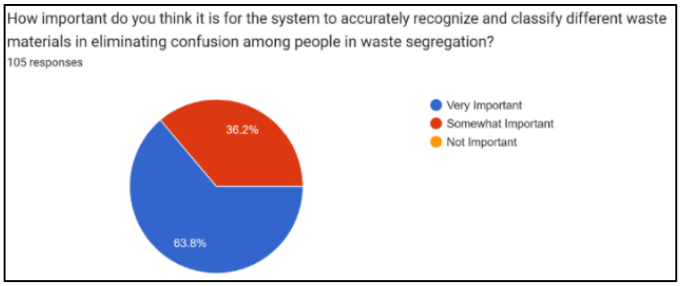


Fig. 6. How important respondents think if the system can recognize and identify waste accurately.

The purpose of asking about the importance of accurate waste material recognition and classification in the proposed system was to gain the respondents' perspective on the significance of leveraging the power of technology in addressing the issue of confusion during waste segregation. As shown in Fig. 6, 63.8% consider it "Very important", which underscores a strong consensus on the important role of technology in enhancing waste management practices. This result suggests a recognition among respondents that an AI-powered waste classification system capable of accurately classifying different waste materials could contribute to eliminating confusion during waste segregation. The positive response toward the importance of technology in waste management aligns with the objectives of the proposed system.

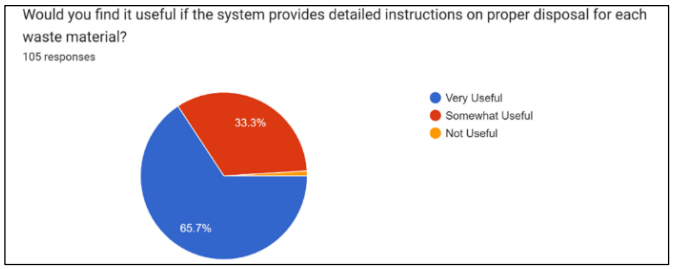


Fig. 7. How useful the system can provide instructions on waste disposal.

The purpose of including the above question in the questionnaire is to understand the users' expectations and preferences regarding the features of the proposed waste classification system. As shown in Fig. 7, the majority of respondents, 65.7%, find it "Very Useful", indicating a strong desire among respondents for a system that not only classifies waste materials but also provides comprehensive guidance on proper disposal. This result suggests that users value not only the identification aspect but also practical guidance on waste management. The demand for detailed instructions aligns with the system's goal of promoting responsible waste disposal practices. It reinforces the importance of integrating informative features within the proposed system to enhance user experience and contribute to more effective waste management habits.

B. Phase 2: CRISP-DM Framework

Following the CRISP-DM framework, the authors collected data from primary and secondary sources [21] to ensure the trained model directly reflects Malaysia's waste context. A total of 12 waste classes are labeled as shown in Fig. 8.

Fig. 8. Sample waste image from dataset for CNN model training.

Fig. 8 shows some sample waste image datasets used for CNN training. Before training, the authors removed the corrupted and low-resolution waste images to maintain the data quality. The images are pre-processed by utilizing data augmentation techniques, including rotation and flipping to improve its ability to recognize more waste variations.

C. Phase 3: System Design

The system is designed by using one of the famous UML diagrams, a use case diagram as shown in Fig. 9 to understand how each component in the system works with each other. By using a use case, the authors can easily figure out how users interact with a system. Use cases help the authors understand the system's functionality, behavior, and interactions in a structured and understandable form. They are valuable for guiding the design, development, and testing of software systems, ensuring that they meet the needs and expectations of users.

D. Phase 4: System Implementation

The authors utilized the power of Jupyter Notebook, Python Programming Language, TensorFlow, and Keras libraries to build and train the deep learning model, Convolutional Neural Network. The authors tried both methods in training the CNN model: train from scratch and train using a pre-trained model. The results from these two methods have shown a large difference.

Fig. 9. System design.

By training the CNN model from scratch, the model achieves a training accuracy of 83.82%, and a validation accuracy of 68.49% (see Fig. 10(a)). In contrast, by using a pre-trained model, which is Visual Geometry Group 16 (VGG16) as the base architecture, it achieves a training accuracy of 80.66% and an increased validation accuracy of 77.62% in waste image classification (see Fig. 10(b)). This further proves the efficacy and performance of using a pre-trained model as a base architecture, as highlighted in previous literature reviews [12] [18].

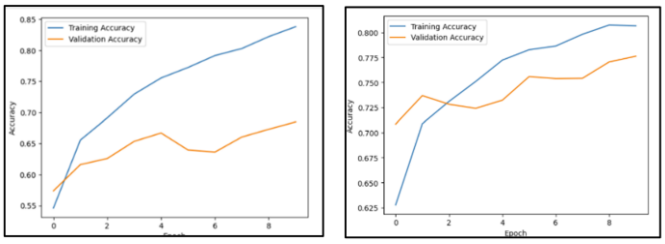


Fig. 10. (a) and (b) show the accuracy of two different methods for training the CNN model.

Transitioning to the system development, the authors used a combination of VS Code, PHP, HTML, CSS, and JavaScript to craft the user interface and system functionalities. What ties everything together is the Flask. By leveraging the Flask framework, the authors successfully integrate the trained CNN model with higher validation accuracy into the system, enabling the user to capture or upload waste images, thus getting immediate waste classification results. At this moment, users can choose to view the customized waste disposal guidelines based on their preference.

E. Phase 5: System Testing

Two main tests were conducted: unit testing and usability testing. The overall unit testing achieved a high pass rate, meaning that the majority of test cases passed successfully. Those test cases that failed will need further investigation and improvement. Some key observations are concluded. Firstly, the CNN model can classify the waste moderately well, but still, some misclassifications were detected, particularly in cases that involve complex waste images.

The result from the usability testing shows that the system consistently provides accurate waste classification results and value disposal guidance to users. Many of them expressed appreciation for the user-friendly interface for completing the tasks given. They are impressed by the power of deep learning in classifying waste images and the sustainable disposal tips provided by the system.

IV. Result and Discussion

This section presents the findings from the development and testing of the proposed waste classification system. The authors begin with system development results by displaying the interfaces, highlighting its design elements that enhance usability and facilitate efficient waste management. Next, the authors present the results of the waste classification testing, which is the main core of the system to showcase the system's accuracy and performance metrics to demonstrate its effectiveness in identifying various waste types. Finally, the authors compare the proposed system with existing solutions, illustrating its advantages in terms of classification capabilities and overall user experience. Through this discussion, the authors aim to provide a comprehensive understanding of the system's contributions to waste management practices.

A. System Development Result

The developed system consists of several features: an informative homepage, waste capture or upload, view classification results and customized disposal guidelines, locate recycling centers, and a dashboard as shown in Fig. 11, Fig. 12, Fig. 13, Fig.14 and Fig.15.

Fig. 11. Informative home page.

Fig. 11 illustrates the system's home page, which provides various useful educational resources and information. This page is designed to help users access essential content related to waste management and recycling.

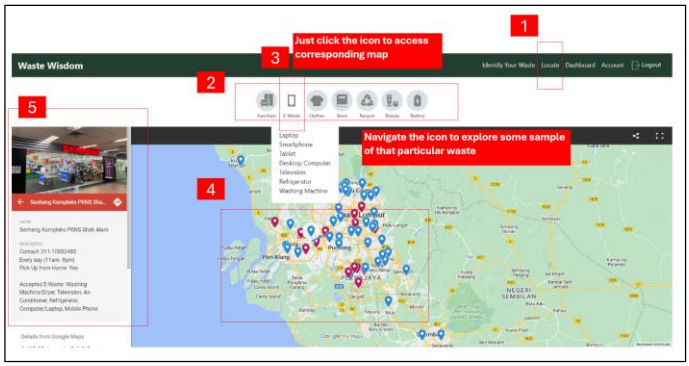


Fig. 12. Locate recycling centers.

Fig. 12 illustrates the system page where it displays the recycling center's location of each different recyclable item. Users can easily navigate through a row of clickable icons representing various items for recycling. For instance, if a user wishes to recycle furniture, they can simply click the furniture icon, and the corresponding map will be displayed.

Fig. 13. Dashboard.

Fig. 13 illustrates the dashboard, including global waste statistic, Malaysia e-waste collection centers by states, Malaysia household waste composition, recycling rate by country and Malaysia landfill status.

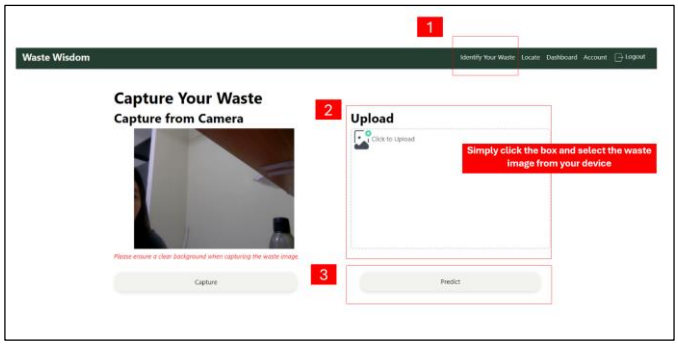


Fig. 14. Capture waste image.

Fig. 14 illustrates the system page where the user can either capture or upload the waste image.

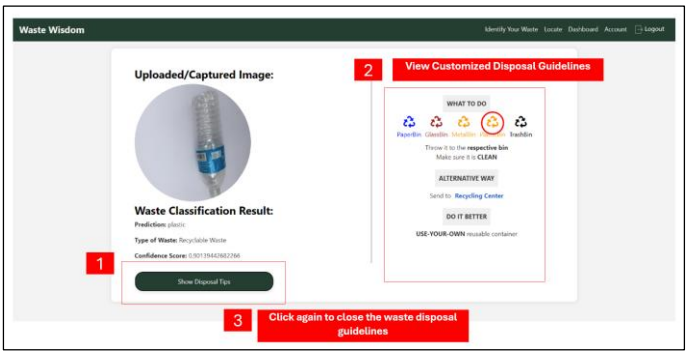


Fig. 15. Waste classification result and guidelines.

Fig. 15 illustrates the system page where the waste classification result and confidence score are displayed. Users can click the "Show Disposal Tips" button to access guidelines for classified waste.

B. Waste Classification Testing Result

This segment presents the results of the waste classification testing, detailing the accuracy and reliability of the system. The findings demonstrate the effectiveness of the classification algorithm, highlighting how well the system identifies various waste types and the confidence scores associated with each prediction.

Fig. 16. Waste items from various categories for testing.

Table I shows the predicted classifications, along with their scores, and indicates whether the model's predictions are considered a pass or fail based on the 60% threshold. The results were derived from the various waste items displayed in Fig. 16. The model correctly classifies 11 out of 12 items, showing that it performs well across a variety of waste types. Even when confidence is lower, it still makes the right prediction in most cases.

TABLE I. WASTE CLASSIFICATION TESTING RESULT

No	Expected	Predicted	Score	Pass/Fail
1	cardboard	cardboard	0.6518	Pass
2	paper	paper	0.5376	Pass
3	glass	glass	0.9991	Pass
4	organic	organic	0.6804	Pass
5	metal	metal	0.9949	Pass
6	plastic	plastic	0.8644	Pass
7	charger	glass	0.6522	Fail
8	mouse	mouse	0.5518	Pass
9	smartphone	smartphone	0.9717	Pass
10	battery	battery	0.9854	Pass
11	laptop	laptop	0.8799	Pass
12	trash	trash	0.4855	Pass

The model performed exceptionally well for categories such as Glass (0.9991), Metal (0.9949), Battery (0.9854), and Smartphone (0.9717). These high scores indicate that the model is highly confident in these predictions, likely due to clear and distinctive features that help distinguish these waste types. Predictions such as Cardboard (0.6518), Organic (0.6804), and Plastic (0.8644) also passed, with confidence scores above the 60% threshold. Although the confidence for these items is slightly lower, the model still correctly identified the waste types, demonstrating good performance overall. On the other hand, the predictions for Paper (0.5376), Mouse (0.5518), and Trash (0.4855), while below the 60% threshold, were correctly classified. The model successfully identified the correct waste category despite its lower confidence. This suggests that even in cases where the model is less sure, it can still make accurate classifications, which is a positive sign for real-world applications.

The only failure in the table is the misclassification of "Charger" as "Glass", with a confidence score of 0.6522. This is possibly due to poor image quality. Another reason might be the image itself's complexity which makes the system confused and misclassifies the waste category. Although the confidence score exceeds the threshold, the predicted output was incorrect. This indicates a specific issue in differentiating Charger from Glass. Therefore, it is important to ensure that the background is clear when capturing or uploading the waste image and further fine-tuning is needed to help the model better differentiate chargers from other materials. But overall, the trained CNN model performed reasonably well in recognizing and classifying common waste.

C. Comparison of Existing and Proposed Systems

This segment compares the proposed system with existing waste classification systems to illustrate the advantages of the proposed system. By comparing important aspects such as key features, model used, dataset, accuracy and user experience, the authors highlight how the proposed system offers enhanced capabilities and improved recycling efficiency, addressing the limitations of current solutions.

Table II shows a clear comparison of key features and models used across several waste classification systems, including DeepWaste [19], EcoScan [20], 101 Trash [22] and proposed system. The proposed system uses a larger and more diverse dataset, specifically tailored to the Malaysian waste context. Since waste generation patterns vary across countries due to cultural differences, this localized focus allows the system to better reflect the types of waste commonly found in Malaysia, enhancing its relevance and accuracy for the target region.

TABLE II. COMPARISON BETWEEN EXISTING SYSTEM AND PROPOSED SYSTEM

Aspects	Deep Waste	Eco Scan	101 Trash	Proposed System
Require Login	X	X	X	/
Model Used	Deep Learning	Not publicly disclosed	X	Deep Learning (CNN – VGG16)
Dataset Used	More than 1,200 waste images	Not publicly disclosed	101 types of trash	Primary waste images source and garbage waste dataset (15,497 images) from Kaggle
Accuracy	88.1%	X	X	77.62%
Capture Waste Image	/	/	X	/
Classification Feature	/	/	X	/
Confidence Score	/	X	X	/
Waste Disposal Guidelines Feature	X	X	/	/
Access Dashboard Feature	X	X	X	/
Locate Feature	X	X	/	/
Country-Based	X	X	/	/
Platform	Mobile	Mobile	Website	Website

[a.] / indicates the feature is present; **X** indicates the feature is absent

Although the proposed system has an accuracy of 77.62%, which is lower than Deep Waste's 88.1%, it uses the VGG-16 deep learning model. This choice provides a strong foundation for improving accuracy with further adjustments and tuning.

In terms of the classification feature, DeepWaste classifies the waste into three categories: Recycle, Trash and Compost [19], whereas EcoScan only differentiates between two categories: Paper and Glass [20]. This demonstrates the proposed system's ability to offer a more complete waste classification approach to enhance the waste segregation practice. It is believed that with more classification options, the system can improve recycling efficiency and support better waste disposal practices.

Key features such as confidence scoring, classification types, and waste disposal guidelines make the proposed system competitive and user-friendly. It also includes a dashboard and location-based functionality, ensuring its utility in local waste management. While its accuracy can be

enhanced, the proposed system's transparency, dataset size, and feature-rich design indicate its potential to outperform existing solutions with further development.

V. Conclusion and Future Work

In conclusion, the developed system is considered to reasonably effectively address the challenges outlined in the problem statement by providing a practical solution to waste classification and waste disposal technologically. The first development of this system can accurately classify common household waste.

By acting as an educational platform, the developed system aims to eliminate people's confusion about waste segregation and educate them about sustainable waste management practices, contributing to a greener environment and, at the same time promoting SDG 12 and SDG 13. By consistently using the system, it is believed that the user can influenced by the system gradually as it provides different essential knowledge about recycling practices and provides details about where they can recycle their specific waste items. Also, the dashboard info might act as an important element to raise people's awareness to start taking action to reduce waste.

In addition, future works would be done by regularly updating the CNN model training dataset and expanding the application to mobile-based. Other than that, rather than focusing on image classification tasks, waste object detection can be taken into future consideration to make it more appealing and allow for real-time processing.

Acknowledgment

I would like to express my heartfelt gratitude to my supervisor, Ms. Chong Fong Kim for her invaluable guidance, patience, encouragement, and insightful feedback during the research process. Special thanks to the Faculty of Data Science and Information Technology (FDSIT) for their support in my academic journey. Furthermore, thank you to my family and friends for their support and collaboration, which made this journey enjoyable. Lastly, I also extend my appreciation to all my educators and external contributors for their dedication to teaching and unwavering commitment to student success. Their collective wisdom and support have profoundly impacted my academic and professional journey, and for that, I am truly grateful.

References

[1] Kumar, A. and Agrawal, A. (2020). Recent trends in solid waste management status, challenges, and potential for the future Indian cities – A review. Current Research in Environmental Sustainability, 2, p.100011. doi:https://doi.org/10.1016/j.crsust.2020.100011.

[2] Yu, P.L., Ab Ghafar, N., Adam, M. and Goh, H.C. (2022). Understanding the Human Dimensions of Recycling and Source Separation Practices at the Household Level: An Evidence in Perak, Malaysia. Sustainability, 14(13), p.8023. doi:https://doi.org/10.3390/su14138023.

[3] Shakil, N.S.M., Azhar, N.A.Z.M. and Othman, N. (2023). Solid Waste Management in Malaysia: An overview. Information Management and Business Review (ISSN 2220-3796), 15(1), pp.86–93.

[4] Razali, F., Daud, D., Weng-Wai, C. and Anthony Jiram, W.R. (2020). Waste separation at source behaviour among Malaysian households: The Theory of Planned Behaviour with moral norm. Journal of Cleaner Production, 271, p.122025. doi:https://doi.org/10.1016/j.jclepro.2020.122025.

[5] Law, J., Lye, C.-T. and Ng, T.-H. (2023). Can environmental literacy and integrated behavioral factors encourage green practices at home? Evidence from Malaysia. Cleaner and responsible consumption, pp.100134–100134. doi:https://doi.org/10.1016/j.clrc.2023.100134.

[6] Abu Bakar, I.A., Ramayan, S.K. and Sangaran, V.S. (2020). Public Inclination Towards Waste Segregation Programme (A Case Study on the Effectiveness Of #Asingkan Campaign in Klang Valley). Idealogy Journal, 5(2), pp.185–198. doi:https://doi.org/10.24191/idealogy.v5i2.241.

[7] Cheng, K.M., Tan, J.Y., Wong, S.Y., Koo, A.C. and Amir Sharji, E. (2022). A Review of Future Household Waste Management for Sustainable Environment in Malaysian Cities. Sustainability, 14(11), p.6517. doi:https://doi.org/10.3390/su14116517.

[8] Chen, B. and Lee, J. (2020). Household waste separation intention and the importance of public policy. International Trade, Politics and Development, 4(1), pp.61–79. doi:https://doi.org/10.1108/itpd-03-2020-0008.

[9] Shakil, N.S.M., Azhar, N.A.Z.M. and Othman, N. (2023). Solid Waste Management in Malaysia: An overview. Information Management and Business Review (ISSN 2220-3796), 15(1), pp.86–93.

[10] Tang, Y.Y., Tang, K.H.D., Maharjan, A.K., Abdul Aziz, A. and Bunrith, S. (2021). Malaysia Moving Towards a Sustainability Municipal Waste Management. Industrial and Domestic Waste Management, 1(1), pp.26–40. doi:https://doi.org/10.53623/idwm.v1i1.51.

[11] Ahmad, I., Chelliapan, S., Abdullah, N. and Ahmad, M.D. (2019). Sanitary Landfill is a Solution in Solid Waste Management or a Silent Threat to Environment: Malaysian Scenario. Open International Journal of Informatics (OIJI), 7, pp.135–146.

[12] Shahab, S., Anjum, M. and Umar, M.S. (2022). Deep Learning Applications in Solid Waste Management: A Deep Literature Review. (IJACSA) International Journal of Advanced Computer Science and Applications, 13(3), pp.381–395.

[13] Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M. and Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data, [online] 8(1). Available at: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8.

[14] Kufel, J., Bargieł-Łączek, K., Kocot, S., Koźlik, M., Bartnikowska, W., Janik, M., Czogalik, Ł., Dudek, P., Magiera, M., Lis, A., Paszkiewicz, I., Nawrat, Z., Cebula, M. and Gruszczyńska, K. (2023). What Is Machine Learning, Artificial Neural Networks and Deep Learning?—Examples of Practical Applications in Medicine. Diagnostics, [online] 13(15), p.2582. doi:https://doi.org/10.3390/diagnostics13152582

[15] Surono S., Rivaldi M., Dewi D.A., Irsalinda N. (2023). New Approach to Image Segmentation: U-Net Convolutional Network for Multiresolution CT Image Lung Segmentation. Emerging Science Journal. Vol 7, No 2 (2023). doi: https://www.ijournalse.org/index.php/ESJ/article/view/1444/pdf

[16] Vo, A.H., Hoang Son, L., Vo, M.T. and Le, T. (2019). A Novel Framework for Trash Classification Using Deep Transfer Learning. IEEE Access, 7, pp.178631–178639. doi:https://doi.org/10.1109/access.2019.2959033

[17] Ruiz, V., Sanchez, A., F. Velez, J. and Raducanu, B. (2019). Automatic Image-Based Waste Classification.

[18] Albashish, D., Al-Sayyed, R., Abdullah, A., Ryalat, M.H. and Almansour, N.A. (2021). Deep CNN Model based on VGG16 for Breast Cancer Classification. 2021 International Conference on Information Technology (ICIT). doi:https://doi.org/10.1109/ICIT52682.2021.9491631.

[19] Narayan, Y. (2021). DeepWaste: Applying Deep Learning to Waste Classification for a Sustainable Planet. 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

[20] EcoScan (2019). EcoScan - Recycle Waste and Get Rewarded. [online] EcoScan app. Available at: https://www.ecoscanapp.eu/.

[21] Mohamed, M. (2020). Garbage Classification (12 classes). [online] www.kaggle.com. Available at: https://www.kaggle.com/datasets/mostafaabla/garbage-classification.

[22] Zero Waste Malaysia (2021). 101 Trash. [online] trashpedia.zerowastemalaysia.org. Available at: https://trashpedia.zerowastemalaysia.org/en/101trash/.

Attention Mechanism-Based CNN-LSTM for Abusive Comments Detection and Classification in Social Media Text

BalaAnand Muthu First[1], Kogilavani Shanmugavadive[2], Veerappampalayam Easwaramoorthy Sathishkumar[3], Muthukumaran Maruthappa[4], Malliga Subramanian[5], Rajermani Thinakaran[6]
Dept. of Computer Science & Engineering, Tagore Institute of Engineering and Technology, Deviyakurichi, Attur, Salem, India[1]
Dept. of Artificial Intelligence, Kongu Engineering College, Perundurai, Erode, Tamil Nadu, India[2]
Dept. of Computing and Information Systems, Sunway University, Selangor, Malaysia[3, 4]
Dept. of Computer Science and Engineering, Kongu Engineering College, Perundurai, Erode, Tamil Nadu, India[5]
Faculty of Data Science and Information Technology, INTI International University, Negeri Sembilan, Malaysia[6]

***Abstract*—Human contact with one another through social networks, blogs, forums, and online news portals and communication has dramatically increased in recent years. People use these platforms to express their feelings, but sometimes hateful comments are also spread. When abusive language is used in online comments to attack individuals such as celebrities, politicians, and products, as well as groups of people associated with a given country, age, or religion, cyberbullying begins. Due to the ever-growing number of messages, it is challenging to manually recognize these abusive comments on social media platforms. This research work concentrates on a novel attention mechanism-based hybrid Convolutional Neural Network - Long Short Term Memory (CNN-LSTM) model to detect abusive comments by getting more contextual information from individual sentences. The proposed attention mechanism-based hybrid CNN-LSTM model is compared with various models on the dataset provided by the shared task on Abusive Comment Detection in Tamil – ACL 2022 which contains 9 class labels such as Misandry, Counter-speech, Xenophobia, Misogyny, Hope-speech, Homophobia, Transphobic, Not-Tamil and None-of-the-above. We obtained an accuracy of 67.14%, 68.92%, 65.35% and 68.75% on Naïve Bayes, Support Vector Machine, Logistic Regression and Random Forest respectively. Furthermore, we applied the same dataset to deep learning models like Convolutional Neural Networks (CNN), Long Short Term Memory (LSTM), Bidirectional-Long Short Term Memory (Bi-LSTM) and obtained an accuracy of 70.28%, 71.67% and 69.45%, respectively. To obtain more contextual information semantically a novel attention mechanism is applied to the hybrid CNN-LSTM model and obtained an accuracy of 75.98% which is an improvement over all the developed models as a process innovation.**

Keywords*—Attention mechanism; hybrid CNN-LSTM model; machine learning model; deep learning model; abusive comments detection

I. Introduction

The rise in web and social media interactions leads to a massive amount of information. Due to the freedom to convey everyone's opinion, sometimes the content posted on Facebook, Twitter, and YouTube may be offensive in nature [1]. Chakravarthi et al., [2] looked into the methods for recognizing several forms of abusive content, including aggressiveness, cyberbullying, hate speech, offensive language, abusive remarks, and abusive comments. Zampieri et al., [3] talked about automating the technique for detecting offensive language. The algorithms are trained using postings that have had the presence of any abusive or objectionable content noted. The entire problem was modeled as a supervised learning problem. In general, the Offensive Language Identification problem can be broadly categorized into Aggression Identification, Abusive Comments Detection, Hate Speech Identification, Offensive Language Identification and Toxic Comments Identification. The Aggression and Hate Speech Identification System classifies the given comments into Non-Aggressive, Covertly Aggressive and Overtly Aggressive. In the Abusive Comments Detection System, the class labels are Xenophobia, Misogyny, Hope-Speech, Homophobia, Transphobic, Not-Tamil, None-of-the-above categories. Toxic comments can be classified into Toxic, Severe Toxic, Obscene, Threat, Insult and Identity Hate categories.

Abuse is the act of making remarks that are hurtful to a particular person or group of people. A phrase that uses vulgar or harsh language in a discourse is referred to as abusive language either in oral form or in text form. The lack of eye contact among users of various social media platforms allows them to speak on the topic fearlessly. Therefore, it is necessary to automatically ban, discourage, or restrict users whose actions are hostile. Online abuse has contributed to issues including low self-esteem, despair, harassment, and in very extreme situations, death. Because abusive content can be communicated in a variety of ways, identifying and handling such comments is crucial and difficult. It is nearly impossible to manually find and remove abusive remarks from a large internet comment stream. Additionally, very few research has been done to identify abusive language in Dravidian languages like Tamil. The goal of this research project is to find instances of abusive language in Tamil-language YouTube comments. At the comment level, each post is tagged with nine different class labels. The data set was taken from the ACL 2022 shared task as in [4] which contains YouTube comments and Twitter posts written in Tamil language. The dataset description along with class labels are shown in Table I.

TABLE I. DATASET DESCRIPTION

Dataset	Comments	Labels
Training	2000	Misandry Counter-speech Xenophobia, Misogyny Hope-speech Homophobia Transphobic Not-Tamil None-of-the-above
Validation	240	
Testing	561	

The term "misandry" refers to established bias towards men. The term "counter-speech" refers to a strategy that presents an alternate story in place of offensive speech to combat hate speech or disinformation. The term "misogyny" refers to inherent bias against women. YouTube comments and posts with the class name "hope-speech" provide encouragement, assurance, advice, inspiration, and insight. The term "xenophobia" refers to a strong dislike of foreigners. The term "homophobia" refers to an antipathy, prejudice, or fear of homosexuals or homosexuality. The term "transphobia" is used to describe a broad spectrum of unfavorable attitudes, sentiments, or behaviors toward transgender persons. Sample comments in Tamil and their respective classes are represented in Table II.

TABLE II. COMMENTS AND CLASSES

Comments	Classes
பேச்சை எதிர்பார்த்தேன். நல்ல விளக்கம் அருமை. நியாயமான முறையில் பதிவு செய்துள்ளீர்கள். நீண்ட காலம் வாழ்வதற்கு வாழ்த்துக்கள். I was looking forward to your speech. Good description Awesome. You have registered reasonably. Congratulationson living a long life.	Hope-speech
இது பெண்மையை பயம் முறுத்தும் புரிதலற்ற பேச்சு.... This is an incomprehensible talk that scares women	Counter Speech
ராஜா நீ நாயி Raja you dog	Homophobia
கரு.நீ பெரியார் கருவா ? Are u egg of Periyar?	Misandry
சொத்து என்பதே இப்போது வந்ததுதான் அதற்கு முன்னமே ஆண்களுக்கு மட்டும் தான் சொத்து Property is what it is now and before that it was only for men	Misogyny
தமிழ் நாடு உனக்கு சோறு போடுது Tamil Nadu will give you rice	None of the above
பிச்சை பையன் எச்ச ராஜா The remnant king of the begging boy	Xenophobia
Seaman is a scumbag. Even a counselor is jealous. Waste talking fellow	Not-Tamil
அலி...நீ உழைக்கிற சாதியா Transgender …you are a working caste	Transphobic

Initial pre-processing steps like tokenization, emoji removal and label encoding are carried out. Significant features are extracted by utilizing Term Frequency-Inverse Document Frequency (TF-IDF), Count Vectorizer methods and fed into machine learning models such as Naïve Bayes (NB), Support Vector Machine (SVM), Logistic Regression (LR) and Random Forest (RF). Word embedding is a technique used by state-of-the-art deep learning models like as Convolutional Neural Networks (CNN), Long Short Term Memory (LSTM) and Bidirectional-Long Short Term Memory (Bi-LSTM) to incorporate the context or high-level meaning of each word present in the comments. Although CNN has been shown to be effective at categorizing text, its classification performance is typically hampered by the frequent omission of essential long-distance sequential data, particularly in sentences with negation and semantic transition. Similar to this, the LSTM model is capable of capturing contextual data, particularly the meaning of long text data. However, this approach is unable to concentrate on the text's most crucial passages. In this work, a hybrid deep learning framework based on an attention mechanism and a combination of CNN and LSTM is presented to identify and categorize abusive remarks in social media text. To precisely identify the abusive comments, the ACL2022 shared task dataset which contains comments in Tamil language sentences was selected as the input of the neural network. CNN layers are used in the proposed hybrid CNN-LSTM model to capture features that specify the word's local information in its context. The text is then modeled with attention signals and sequence learning is performed using LSTM layers. Finally, the input from earlier neural networks is combined and offensive remarks are detected using fully linked networks. The impact of applying feature extraction and word embedding techniques to various learned models is also examined in this research work.

All of these cutting-edge models are compared to the suggested innovative attention mechanism-based hybrid CNN-LSTM model's performance. The models must categorize the comments into one of the nine classes when the test data is given. The objectives of this research work are 1) Abusive comments detection from YouTube comments in Tamil using machine learning and deep learning models; and 2) Embedding an attention mechanism in a hybrid CNN-LSTM model.

The research work that is currently available and the tools created for abusive comment detection are described in Section II. The proposed system models are mentioned in Section III. The performance evaluation with models' comparison is represented in Section IV and Section V ends with contributions and prospects for future direction.

II. LITERATURE REVIEW

Devlin et al., [5] described an approach of detecting emotions in social media comments in various languages. The

authors employed Neural Network (NN) layers and adoptive parameters to produce different models. Sanh et al. [6] show that using information distillation, a Bidirectional Encoder Representation from Transformers (BERT) model can be made 40% smaller while preserving 97% of its language knowledge.

Vinay et al., [7] classified the comments as personal attacks and not personal attacks. They employed LSTM with word embedding, CNN with word embedding, and CNN with character embedding. Out of all the models, CNN with character-level embedding achieves the best results. A hybrid method that applies sentiment analysis to machine learning approaches was developed by Hasan et al., [8]. In their study, supervised techniques like NB and SVM were in the investigation of strategic positions.

Huang et al., [9] discovered that hierarchical LSTMs leave extensive context modeling, by allowing them to enhance sentiment categorization significantly. They used LSTMs in particular because they tackle the aforementioned vanishing gradient problem. Zhao et al. [10] used CNNs to test various feature embeddings, ranging from character to sentence level. The researchers expected that the word embedding layer would be critical for sentiment analysis since short texts have a limited quantity of contextual information. They discovered that character-level embedding worked better than other embeddings on one dataset and behaved similarly well on the other using the two datasets they utilized.

Pinkesh et al., [11] classified a tweet as racist, sexist or neither. They conduct extensive research with several deep learning architectures to learn semantic embedding. They tested on a 16K annotated tweets benchmark dataset and revealed that deep learning approaches beat machine learning algorithms. Rotaru et al., [12] discovered how essential phonetic symbols affect people's emotions. Their model's phonetic transcription feature maps show that their model performs fantastically in texts that are overly dense and filled with incorrect words. Additionally, they think that including this grammatical precedence in the sub-word design will improve the model's performance.

Sharif et al., [13] recognized aggressive texts in Bengali using a weighted ensemble strategy and developed methods such as m-BERT, Distill-BERT, Bangla-BERT, and XLM-R. These mentioned models work better than regular methods. Shreelakshmi et al., [14] worked with messages that include code-mixed data from Hindi and English that originate from multilingual consumers. Many previous methodologies omitted data from these low-resource languages.

Sentiment polarity of blended text was utilized by Sevda et al., [15] to classify sentences into those with three emotional meanings. According to this paper, shared parameters are used to translate words. They also showed a simple pre-processing technique based on clustering for gathering variations in code-mixed text. Raut et al. [16] investigated methods for extracting data from Twitter tweets. They also studied supervised learning techniques that could be utilized to identify textual tweet polarity, including SVM for Document Classification. Their conclusion suggests that SVM can recognize text features like a big feature set, or a sparse instance vector.

Cambria et al., [17] discussed new avenues for performing sentiment analysis. During the initial stage itself, a set of emotion class labels were derived from 5,553 tweets using inductive coding. Malliga et al., [18] presented an offensive language detection system using an adapter and transformer-based model. The sentimental analysis of Twitter data was carried out by Barbosa et al., [19] utilizing machine learning and deep learning techniques. They gathered test data from Twitter and examined tweets using syntactic components such as symbols, re-tweets, emoticons, tags, links, punctuation, and exclamation points. They have employed the polarity classifier and the subjectivity classifier as their two primary classifier types. A high-quality collection of Hindi-English code-switch data with 15,744 YouTube comments was produced by Ray et al., [20]. They looked at the collection's growth, the findings of trend analysis using the sample as a guide, and the consistency of the annotators.

Mathur et al., [21] developed a Hindi-English code flipped dataset into Abusive hate speech and non-offensive. A multi-label hostility detection dataset was provided by Bhardwaj et al. [22] and was divided into five categories using the BERT algorithm: fake, hate, offensive, defamatory, and non-hostile. Mulki et al., [23] classified the dataset of Twitter into abusive, hate and normal classes using the Naïve Bayes algorithm. Hassan et al., [24] used SVM and CNN+Bi-LSTM to classify offensive Arabic texts and achieved better results for the CNN+Bi-LSTM approach.

Kogilavani et al., [25] carried out sentiment analysis on Tamil code-mixed data. From the dataset, significant features are identified through a hybrid model. Leite et al. [26] reported a toxic language dataset made up of 21000 tweets that were categorized using the BERT algorithm into GBTQ+phobia, racism, insult, xenophobia, obscenity, misogyny, and non-toxic language. The brief literature review helped to realize the need to identify abusive comments that are present in the Tamil language.

Ashraf et al., [27] presented a method to detect the contextual information present in YouTube comments based on topics such as politics, religion and others. The context is identified with the help of linguistic information present in the comments. Jahan et al., [28] employed a transliterated code-mixed dataset for abusive comments detection in Bangla-English data. The result shows that the transliterated dataset does not improve the accuracy of the system.

III. Materials and Methods

Fig. 1 represents the workflow of the methodologies used in this proposed work. As the first step, pre-processing such as tokenization, emoji removal and label encoding are carried out for the dataset.

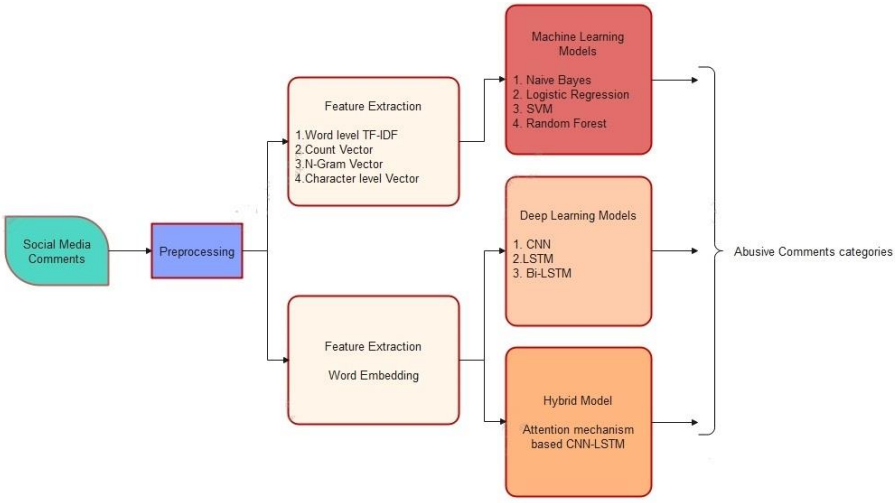


Fig. 1. Proposed system work flow.

Significant features are extracted from Tamil sentences by utilizing TF-IDF and Count Vectorizer methods. The resultant word vectors are fed into machine learning models such as NB, SVM, LR and RF. Word embedding is a technique used by deep learning models like as CNN, LSTM, and Bi-LSTM to incorporate the context or high-level meaning of each word present in the comments. Traditional approaches represent text in terms of TF, IDF, TF-IDF with sparse features. Recently deep learning models have been adopted with word embedding to extract contextual information from the text data. However, these deep learning models do not consider sequential information among the sentences. LSTM is used to capture the semantics of lengthy texts. However, this methodology is unable to focus on important textual passages. When used with text data, this restriction lowers the accuracy of deep learning models. It is necessary to concentrate on particular passages of the text, which are very similar to how people read. The goal of this research work is to suggest a hybrid CNN-LSTM model that is attention-based and can identify offensive comments. The suggested method uses a convolution layer as its first step in capturing attention signals, which reflect the local information of each word in its context. The text is then modeled using LSTM using attention signals. A word with a larger attention weight typically refers to more valuable information.

A. Feature Extraction

The process of extracting features from text input and generating numerical representations of them so that learning models can use them to produce predictions is known as feature extraction. The words or concepts that appear most frequently and have the fewest relevant data start to take precedence over the ones that appear less frequently when a vector is constructed from a text using word frequency. To avoid rating terms in papers purely on their frequency in a single text, as suggested by Sajeetha et al., [29] word frequency must be rescaled.

1) TF-IDF: With TF-IDF, words are given varying weights based on how important they are to the document. It calls attention to a particular problem that, while uncommon in our corpus, is quite crucial. Textual data has therefore been converted into real-valued vectors using TF-IDF. How frequently a word appears in a publication is referred to as term frequency. It might be compared to the likelihood of a term appearing in a document. It establishes the proportion between the frequency of a word (wi) and the overall number of words in the comment (cj). The inverse document frequency statistic can be used to assess how frequently or infrequently a phrase appears across all the documents in a corpus. By dividing the entire number of comments in the dataset by the number of comments that contain the term t, it is possible to get the logarithm of the overall term.

2) Count vector: The Count Vectorizer is used to turn a text into a vector based on how frequently each word appears in the text as a whole. This is helpful when organizing each word in a vector and working with numerous texts of this type. Each distinct word is represented by a column in a matrix created by Count Vectorizer, and each sample of text from a remark is represented by a row. The value of each field indicates how many phrases were contained in the text sample.

3) N-Gram vector: The continuous or neighboring sequence of words in a sentence is represented using the N-gram vector. After performing tokenization and stop words removal process, N-Grams from the text can be represented in vector format.

4) Character level vector: To generate a character level vector, the input sentences are decomposed into a sequence of characters including special characters. To extract morphological information from the text, this kind of character level vector is used.

5) Word embedding: Word embedding, according to Sajeetha et al., [30] is a kind of word vector that makes it possible to represent words with similar meanings. Individual words are often represented in word embedding as real-valued vectors in a predetermined vector space. Since each word is translated into a single vector and its values are learned like that of a neural network, the method is typically categorized under deep learning models. The main concept of the approach is to represent each word with dense sparse representations. A real-

valued vector with a sizable number of dimensions—often tens or hundreds—represents each word.

B. Machine Learning Models

1) Naive Bayes (NB): A group of classification methods known as NB classifiers are based on Bayes' Theorem, which states that all pairings of feature pairs are conditionally independent given the value of the class variable. Mathematically, Bayes theorem can be stated as in (1).

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)} \quad (1)$$

Where y is the class variable and X is the input feature vector. NB learners and classifiers may be extremely fast in comparison to more complex algorithms. Each distribution can be estimated as a one-dimensional distribution individually since the class conditional feature distributions are separated. As a result, problems brought on by the curse of dimensionality are lessened.

2) Support Vector Machine (SVM): Both classification and regression problems can be resolved using the supervised machine learning method known as SVM. Even if categorization more closely reflects the data, regression has downsides. The SVM technique categorizes n-dimensional space by trying to find the best judgment boundary or line so that the following data points can be quickly assigned to the right category. The best choices are known as hyperplanes. When there are only two input features, the hyperplane is essentially a line. When there are three input features, the hyperplane transforms into a two-dimensional plane. Once a hyperplane is generated, it is used to make predictions or classifications by using hypothesis function *h* as in (2) where *x* is input feature, *w* is weight and *b* is bias.

$$h(x_i) = \begin{cases} +1 \; if \; w.x + b \geq 0 \\ -1 \; if \; w.x + b < 0 \end{cases} \quad (2)$$

3) Logistic Regression (LR): LR is a technique for estimating the probability of a discrete result given an input variable. LR models often yield binary outputs, such as true or false, yes or no, and so forth. It is a simple classification model that produces top-notch results and has linearly separable classes. It is a typical categorization technique used in the industrial sector. When performing classification tasks, the analytical technique of logistic regression is useful for assessing if a new sample fits into a particular category. The following equation as in (3) represents logistic regression where *x* is the input feature, *y* is the output variable, b_0 is bias and b_1 is the coefficient for input.

$$\mathrm{y} = \frac{e^{(b0+b1x)}}{1+e^{(b0+b1x)}}) \quad (3)$$

4) Random Forest (RF): The supervised machine learning algorithm RF is frequently employed to address classification and regression problems. As its name suggests, RF is made up of a sizable number of connected decision trees. The class with the highest score is chosen as our model's projection from among the predictions made by each tree in the random forest. It produces decision trees from a variety of data using regression and majority voting, respectively. One of the key characteristics of the RF algorithm was its capacity to handle data sets with both continuous and categorical variables, as in classification and regression issues. Categorization issues, yield improved outcomes.

C. Deep Learning Models

1) Convolutional Neural Network (CNN): CNN is a deep learning technique built on the Multi-Layer Perceptron (MLP). Features are retrieved and used for classification tasks in conventional algorithms. In CNN, higher concepts are created by a series of convolutional layers once these features are automatically extracted. The three main characteristics of CNN are convolution, pooling, and fully connected layers. The most fundamental part of a CNN is the convolution layer, which consists of a number of separate filters commonly referred to as masks or kernels. The input and filters are convolved to produce either an activation map or a feature map. The pooling layer comes next, and its function is to minimize the input's spatial dimensions. Its objective is to simplify the representation of intricate layers. The pooling layer's output is flattened into a single vector and sent as the fully connected layer's input. Fully linked layers, which are utilized to carry out a particular task, such as classification, are the vectors produced by a multilayer perceptron's several convolutions and pooling processes.

2) Long Short Term Memory (LSTM): LSTMs are one type of Recurrent Neural Network (RNN) that stores intermediate outcomes in LSTM networks. When linking old knowledge to fresh information in a traditional RNN, the problem of vanishing gradients usually arises. There is only one layer in the repeating module of a typical RNN. LSTM holds information in a gated cell. A cell can accumulate information and can read and write into its memory. By default, the LSTM has a propensity to retain information for a long time. The Sequential 3 model has a 20000 feature maximum, 128 embedding dimensions, a 40 sequential length, and a 196 LSTM out. The Spatial dropout1d layer receives the output of the Embedding layer, the LSTM layer receives the output of the LSTM layer, and the dense layer receives the output of the dense layer. Nine nodes in the dense layer function as sigmoid activation nodes. The categorical cross entropy loss function and Adam optimizer are used to train the LSTM.

3) Bidirectional-Long Short Term Memory (Bi-LSTM): The process of building a NN that can store sequence information both future to past and past to future is known as Bi-LSTM. Unlike a standard LSTM, a Bi-LSTM has input those flows in both directions. With a standard LSTM, we may make input flow in one way, either backward or forward. However, there is a way to maintain both the present and the future while allowing information to move both ways. In the suggested system, the values for maximum features, embedding size, sequential length, and lstm out are all set to 20,000. The spatial dropout1d layer receives the output of the embedding

layer, the dense layer receives the output of the bidirectional layer as input, and the bidirectional layer receives the output of the dense layer. Nine nodes in the dense layer employ an activation function known as a softmax. The Nadam optimizer and the categorical cross-entropy loss function are used to train the Bi-LSTM.

D. Attention Mechanism-Based Hybrid Model

By fusing CNN with LSTM, the suggested approach creates a novel hybrid model based on an attention mechanism. Through an attention mechanism, neural networks facilitate the dynamic selection of pertinent features from text data. Either the raw input or its higher level representation can receive it directly. This attention mechanism's calculations involve weighing the order of text pieces and giving more weight to pertinent text information. Mainly attention mechanism is used to capture the context between sentences so that if the sentence does not have direct abusive statements also, will be detected by this mechanism. The proposed attention mechanism-based CNN-LSTM hybrid model structure is represented in Fig. 2.

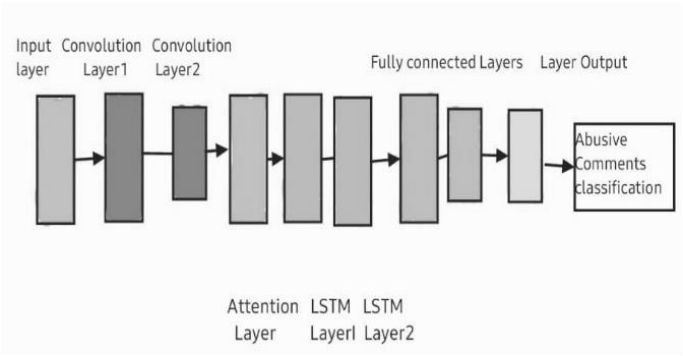


Fig. 2. Attention mechanism-based CNN-LSTM hybrid model.

To develop a CNN+LSTM hybrid model, first CNN layers are added, then attention mechanisms, LSTM layers, and Deep Layers on the output. The first LSTM layer of the LSTM+CNN model receives word vectors for each token in the sequences. This architecture can be seen as defining both the CNN Model for feature extraction and the LSTM Model for feature interpretation over time steps. The Sequential 7 model's maximum features input is set to 20,000, the embedding dimension to 128, the sequential length to 40, and the LSTM out to 196. With a pooling size of 2, a kernel size of 3, and activation functions of ReLU and sigmoid, Conv1d has 128 output filters. Text sequences are created with Conv1d. The output of the embedding layer is provided as input to the conv1d layer, the output of the conv1d layer is input to the MaxPooling1d layer, the output of the MaxPooling1d layer is input to the LSTM layer, and the output of the LSTM layer is input to the dense layer. In the dense layer, nine nodes have sigmoid activation functions. CNN+LSTM is trained using the Adam optimizer and the categorical cross entropy loss function.

IV. Performance Evaluation

Initially, the dataset is applied to word level TF-IDF, count vector, N-Gram vector and character level vector. The proposed system chooses the models based on accuracy metrics instead of loss because the model with minimum loss may not be the model with the best metric. The accuracy obtained by all these feature extraction techniques that are applied to various machine learning models is represented in Table III. The results show that, for all the models, the best accuracy is produced by the character level vector which extracts morphological features from the given text input. Among the entire machine learning models SVM model produces the highest accuracy of 0.6892.

A. Experiments

1) Precision: Precision is the proportion of positive class predictions that fall into that category. Precision values are calculated using as in (4) where TP is True Positive and FP is False Positive.

$$\text{Precision} = TP / (TP + FP) \quad (4)$$

2) Recall: Recall is a metric that indicates how well the model properly detects True Positives. The recall is determined as in (5) where FN is False Negative.

$$\text{Recall} = TP / (TP + FN) \quad (5)$$

3) F1-Score: The weighted mean of Precision and Recall is the F1 Score. This score is generated using an equation as in (6) and takes into consideration both false positives and false negatives.

$$\text{F1 Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision}) \quad (6)$$

The precision, recall, F1-Classification reports of machine learning models such as NB, SVM, LR and RF are presented in Table IV. The result shows that the highest precision of 0.74 is obtained for counter-speech class abusive comments by the NB model. LR produces a precision value of 0.80 for the Not-Tamil type of abusive comments. A precision value of 1.00 is produced by SVM for the Misandry type and by RF for the Not-Tamil type.

TABLE III. Accuracy Comparison of various Machine Learning Models based on Feature Extraction

Models	Word level TF-IDF	Count Vector	N-Gram Vector	Character Level Vector
Naïve Bayes	0.6312	0.6417	0.6315	0.6714
Support Vector Machine	0.6291	0.6010	0.6718	0.6892
Logistic Regression	0.6342	0.6214	0.6171	0.6535
Random Forest	0.6593	0.6432	0.6154	0.6875

TABLE IV. CLASSIFICATION REPORT OF MACHINE LEARNING MODELS

Class Label	Naïve Bayes			Support Vector Machine			Logistic Regression			Random Forest			Support
	P	R	F1-Score	P	R	F1-Score	P	R	F1-Score	P	R	F1-Score	
None-of-the-above	0.00	0.00	0.00	0.25	0.03	0.05	0.00	0.00	0.00	0.00	0.00	0.00	36
Misogyny	0.50	0.12	0.20	0.33	0.12	0.18	0.33	0.12	0.18	0.33	0.12	0.18	8
Misandry	0.00	0.00	0.00	1.00	0.09	0.17	0.67	0.18	0.29	0.75	0.27	0.40	11
Counter-speech	0.74	0.36	0.48	0.69	0.57	0.62	0.73	0.61	0.66	0.64	0.56	0.59	104
Xenophobia	0.00	0.00	0.00	0.40	0.08	0.14	0.50	0.21	0.29	0.50	0.08	0.14	24
Homophobia	0.67	0.98	0.80	0.72	0.97	0.83	0.74	0.96	0.83	0.73	0.96	0.83	346
Hope-Speech	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1
Transphobic	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2
Not-Tamil	0.00	0.00	0.00	0.00	0.00	0.00	0.80	0.14	0.24	1.00	0.07	0.13	29

4) Accuracy: The accuracy of classification models is one of the factors to consider while evaluating them. Informally, accuracy refers to the suggested model's proportion of correct predictions, which is determined as in (7).

$$\text{Accuracy} = (TP+FP) / \text{Total Predictions} \quad (7)$$

Fig. 3 represents the accuracy obtained by NB, LR, RF and SVM learning models. The result shows that the highest accuracy of 68.92 is obtained by SVM for the given dataset.

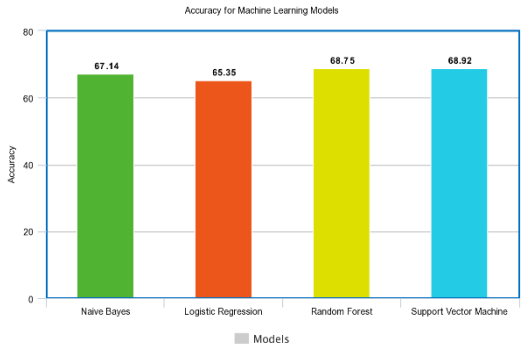


Fig. 3. Performance evaluation of machine learning models based on accuracy.

Fig. 4 denotes the accuracy of CNN, LSTM, Bi-LSTM deep learning models after extracting significant features using word embedding. The result shows that among the three mentioned models, the LSTM deep learning model performs better with an accuracy of 71.67%.

From the deep learning models, the highest accuracy obtained from two models such as CNN and LSTM are combined to generate the hybrid model. The attention mechanism is implemented in between Convolutional layers and LSTM layers in a hybrid model. The highest accuracy machine learning model such as SVM and deep learning model such as LSTM is compared with the hybrid model in Fig. 5. The result shows that the best accuracy of 75.98 is obtained by the attention mechanism-based hybrid CNN-LSTM model.

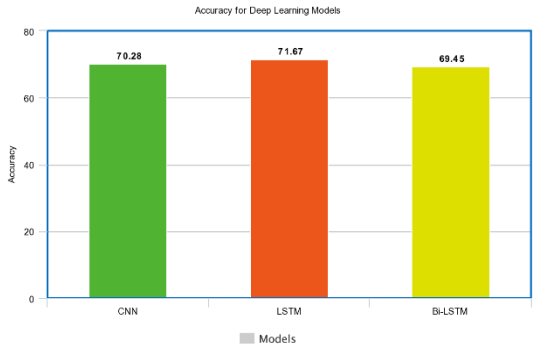


Fig. 4. Performance evaluation of deep learning models based on accuracy.

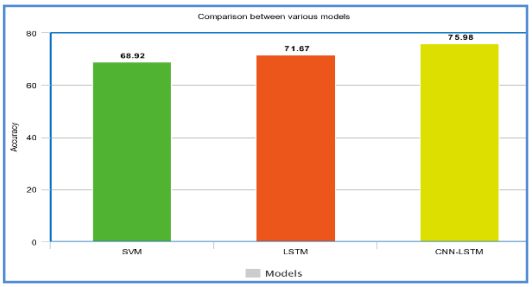
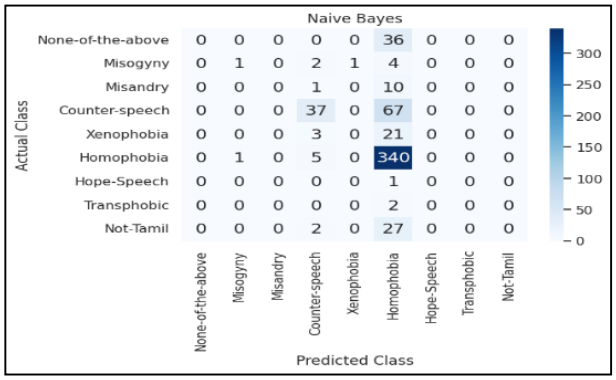
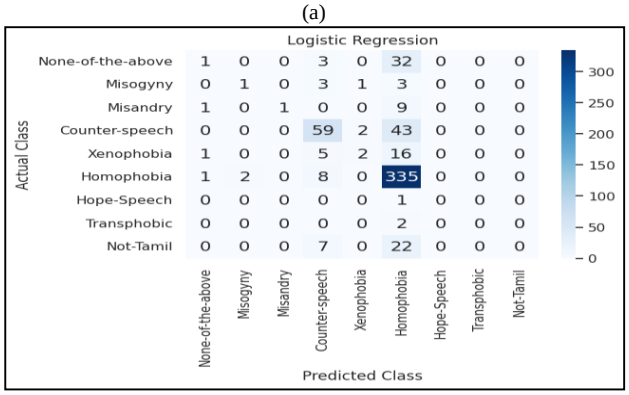


Fig. 5. Comparison of performance evaluation of machine learning, deep learning and attention mechanism-based hybrid CNN-LSTM model.

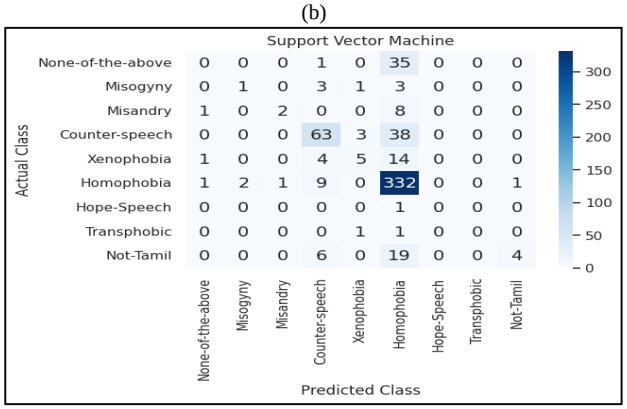
5) Confusion matrix: The confusion matrix is a useful tool for assessing classifiers' abilities to discriminate between data from different classes. In the confusion matrix form, TP and TN are displayed when the classifier is functioning well. FP and FN represent occasions where the classifier is incorrect. A confusion matrix is a table containing the given dataset's 9 classifications and dimensions of 9 x 9. Good accuracy is gained by employing the values along the confusion matrix's diagonal. The confusion matrix for machine learning algorithms like NB, SVM, LR and RF is shown in Fig. 6(a) to 6(d).



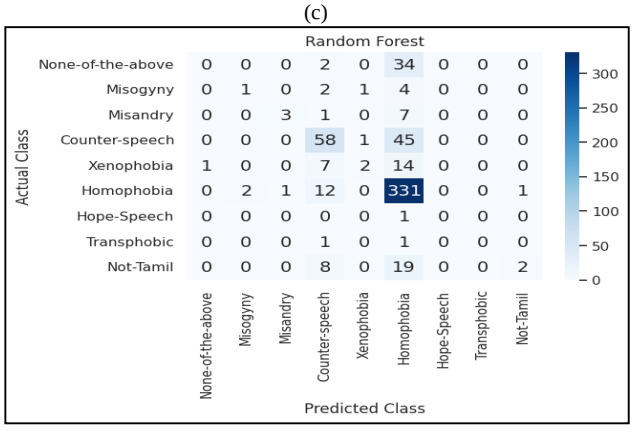
(a)



(b)



(c)

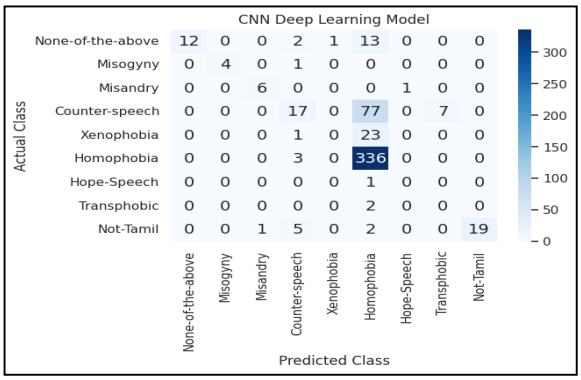


(d)

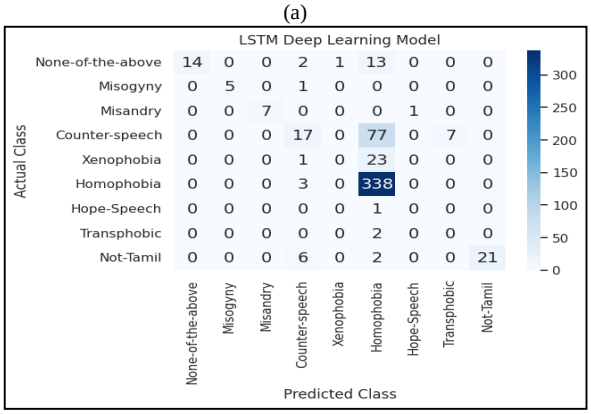
Fig. 6. (a) Confusion matrix - Naïve bayes, (b) Confusion matrix - Logistic regression, (c) Confusion matrix - Support vector machine, (d) Confusion matrix - random forest.

From the confusion matrices in Fig. 6(a) to 6(d), it is understood that out of 346 samples in Homophobia type of abusive comments, the Naïve Bayes models correctly classifies 340 comments, SVM classifies 332 comments, LR classifier 335 comments and RF classifies 331 comments. Fig. 7(a) to 7(c) represents, the confusion matrix obtained by CNN, LSTM and Bi-LSTM deep learning models.

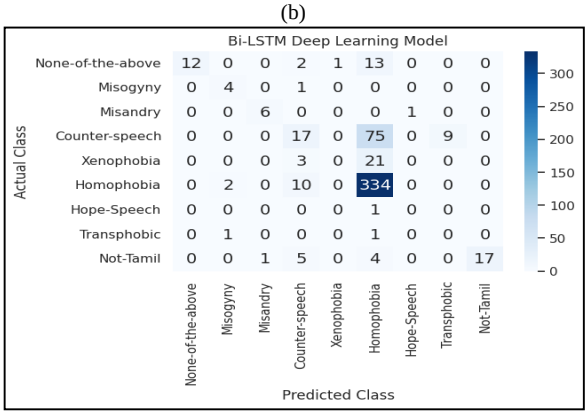
The confusion matrix in Fig. 7(a) to 7(c) shows that out of three deep learning models, the LSTM model outperforms by classifying 402 comments correctly out of 561 comments. Fig. 8 represents the confusion matrix of the proposed attention mechanism-based hybrid CNN-LSTM model. The result shows that 421 comments were correctly classified.



(a)



(b)



(c)

Fig. 7. (a) Confusion matrix - CNN Model, (b) Confusion matrix LSTM Model, (c) Confusion matrix - Bi-LSTM Model.

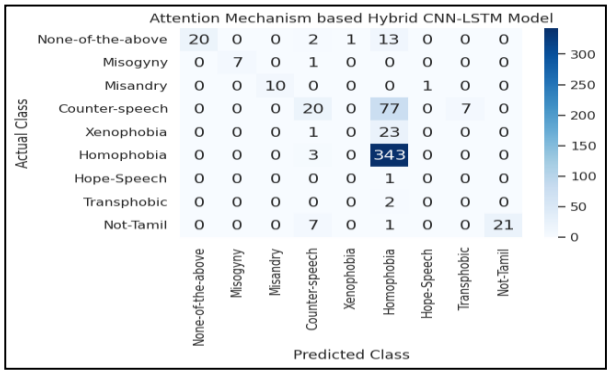


Fig. 8. Confusion matrix - Hybrid CNN-LSTM model.

6) Matthews correlation coefficient: The range of the Matthews Correlation Coefficient (MCC) is [-1,1]. The anticipated values will closely match the actual classification if the value is close to 1, which indicates that the prediction was quite accurate. There is no association between our variables if MCC is equal to 0. Values near -1 imply an inverse relationship between the true and anticipated classes. MCC is calculated using the formula specified as in Eq. (8).

$$MCC = \frac{(TP*TN-FP*FN)}{\sqrt{(TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)}} \quad (8)$$

7) Cohen's Kappa: The concept of calculating the agreement between the Predicted and the True Labels—which are viewed as two random categorical variables—is the foundation of Cohen's Kappa. By creating a confusion matrix and computing the marginal rows and marginal column distributions, it is possible to compare two category variables. Kappa value is calculated using the equation as in (9) where p_o is observed probability and p_e is expected or predicted probability which is calculated from the confusion matrix.

$$Kappa = \left(\frac{p_o-p_e}{1-p_e}\right) \quad (9)$$

8) Error rate: The error rate is calculated by subtracting accuracy from 1 and it is calculated using equation as in (10). For example, the sample sentence "இது பெண்மையை பயம் முறுத்தும் புரிதலற்ற பேச்சு....(This is an incomprehensible talk that scares women ....) is the abusive comment of counter speech category. However, due to the context related to women, it is wrongly classified as a misogyny type.

$$Error\ rate = 1 - accuracy \quad (10)$$

The MCC, kappa value and error rate of all the machine learning models are represented in Table V. Out of all the mentioned machine learning models, SVM obtained the highest MCC value of 0.69 and Kappa value of 0.67. Based on accuracy, the error rate is calculated and it is less for both SVM and Random Forest.

TABLE V. MCC, KAPPA AND ERROR RATE OF MACHINE LEARNING MODELS

Machine Learning Model	MCC	Kappa	Error rate
Naïve Bayes	0.66	0.58	0.32
Support Vector Machine	0.69	0.67	0.31
Logistic Regression	0.64	0.64	0.34
Random Forest	0.67	0.64	0.31

V. CONCLUSION AND FUTURE WORK

In this research work, a hybrid CNN-LSTM model with an attention mechanism has been effectively presented to detect and categorize offensive comments into subcategories including misandry, counter-speech, xenophobia, misogyny, hope-speech, homophobia, transphobia, not-in-Tamil, and none of the above. This research work initially uses a variety of machine learning models, including NB, LR, SVM and RF along with a variety of feature extraction methods, including word level TF IDF, count vector, N-Gram vector, and character level vector. SVM with character level vector based feature extraction method achieves the best accuracy of 68.92%. Word embedding-based deep learning models including CNN, LSTM, and Bi-LSTM were applied to the dataset in order to further increase accuracy. The experimental findings demonstrated that significant features are automatically retrieved, and individual CNN, LSTM models achieve an accuracy of 70.28% and 71.69%, respectively. This result motivates us to develop a hybrid model by combining CNN with LSTM. The suggested work utilizes a novel attention mechanism between CNN and LSTM layers which is called the hybrid CNN-LSTM model to extract more pertinent features from the dataset and achieve superior accuracy of 75.98%. The proposed work detects abusive comments from code-mixed data by applying state-of-the-art machine learning, and deep learning models. Further, the accuracy is improved by applying the proposed attention mechanism-based hybrid CNN-LSTM model. In the future, transfer learning methodology for the identification and classification of abusive remarks may be utilized to improve accuracy further.

REFERENCES

[1] K. Subramaniam, W. Wider, A. Vasudevan, N. Khan, and A. Kohli, "Transitions of value creation from traditional media to social media architecture," Online Journal of Communication and Media Technologies, 13(4), e202356, 2023.

[2] B. R. Chakravarthi, R. Priyadharshini, V. Muralidaran, N. Jose, S. Suryawanshi, E. Sherly, and J. P. McCrae, "Dravidiancodemix: Sentiment analysis and offensive language identification dataset for Dravidian languages in code-mixed text," Language Resources and Evaluation, 56(3), pp. 765-806, 2022.

[3] M. Zampieri, S. Malmasi, P. Nakov, S. Rosenthal, N. Farra, and R. Kumar, "Semeval-2019 task 6: Identifying and categorizing offensive language in social media (offenseval)," arXiv preprint arXiv:1903.08983, 2019.

[4] R. Priyadharshini, B. R. Chakravarthi, S. C. Navaneethakrishnan, T. Durairaj, M. Subramanian, K. Shanmugavadivel, and P. K. Kumaresan, "Findings of the shared task on Abusive Comment Detection in Tamil," In Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages. Association for Computational Linguistics, May 2022.

[5] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.

[6] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," arXiv preprint arXiv:1910.01108, 2019.

[7] V. Singh, A. Varshney, S. S. Akhtar, D. Vijay, and M. Shrivastava, "Aggression detection on social media text using deep neural networks," In Proceedings of the 2nd workshop on abusive language online (ALW2), pp. 43-50, October, 2018.

[8] A. Hasan, S. Moin, A. Karim, and S. Shamshirband, "Machine learning-based sentiment analysis for twitter accounts," Mathematical and computational applications, 23(1), pp. 11, 2018.

[9] M. Huang, Y. Cao, and C. Dong, "Modeling rich contexts for sentiment classification with lstm," arXiv preprint arXiv:1605.01478, 2016.

[10] J. Zhao, S. Mudgal, and Y. Liang, "Generalizing word embeddings using a bag of subwords," arXiv preprint arXiv:1809.04259, 2018.

[11] P. Badjatiya, S. Gupta, M. Gupta, and V. Varma, "Deep learning for hate speech detection in tweets," In Proceedings of the 26th international conference on World Wide Web companion, pp. 759-760, April, 2017.

[12] A. S. Rotaru, and G. Vigliocco, "Constructing semantic models from words, images, and emojis," Cognitive science, 44(4), pp12830, 2020.

[13] O. Sharif, E. Hossain, and M. M. Hoque, "M-bad: A multilabel dataset for detecting aggressive texts and their targets," In Proceedings of the Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situations, pp. 75-85, May 2022.

[14] K. Sreelakshmi, B. Premjith, and K. P. Soman, "Detection of hate speech text in Hindi-English code-mixed data," Procedia Computer Science, 171, pp.737-744, 2020.

[15] S. Mammadli, S. Huseynov, H. Alkaramov, U. Jafarli, U. Suleymanov, and S. Rustamov, "Sentiment polarity detection in Azerbaijani social news articles," In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019), pp. 703-710, September 2019.

[16] V. Shweta Raut, and M. M. Nashipudimath, "Performance Analysis of Feature Selection for Twitter Sentiment Analysis: Classification approach," International Journal of Advanced Research in Computer and Communication Engineering, pp67-74, 2018.

[17] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," IEEE Intelligent systems, 28(2), pp. 15-21, 2013.

[18] M. Subramanian, R. Ponnusamy, S. Benhur, K. Shanmugavadivel, A. Ganesan, D. Ravi, and B. R. Chakravarthi, "Offensive language detection in Tamil YouTube comments by adapters and cross-domain knowledge transfer," Computer Speech & Language, 76, pp. 101404, 2022.

[19] L. Barbosa, and J. Feng, "Robust sentiment detection on twitter from biased and noisy data," In Coling 2010: Posters, pp. 36-44, August 2010.

[20] D. Sitaram, S. Murthy, D. Ray, D. Sharma, and K. Dhar, "Sentiment analysis of mixed language employing Hindi-English code switching," In 2015 International Conference on Machine Learning and Cybernetics (ICMLC), vol. 1, pp. 271-276, IEEE, July D. 2015.

[21] P. Mathur, R. Shah, R. Sawhney, and Mahata, "Detecting offensive tweets in hindi-english code-switched language," In Proceedings Of The Sixth International Workshop On Natural Language Processing For Social Media, pp. 18-26, July 2018.

[22] M. Bhardwaj, M. S. Akhtar, A. Ekbal, A. Das, and T. Chakraborty, "Hostility detection dataset in Hindi," arXiv preprint arXiv:2011.03588, 2020.

[23] H. Mulki, H. Haddad, C. B. Ali, and H. Alshabani, "L-hsab: A levantine twitter dataset for hate speech and abusive language," In Proceedings of the Third Workshop on Abusive Language Online, pp. 111-118, August 2019.

[24] A. Hasan, S. Moin, A. Karim, and S. Shamshirband, "Machine learning-based sentiment analysis for twitter accounts," Mathematical and Computational Applications, 23(1), pp. 11, 2018.

[25] K. Shanmugavadivel, S. H. Sampath, P. Nandhakumar, P. Mahalingam, M. Subramanian, P. K. Kumaresan, and R. Priyadharshini, "An analysis of machine learning models for sentiment analysis of Tamil code-mixed data," Computer Speech & Language, 76, pp. 101407, 2022.

[26] J. A. Leite, D. F. Silva, K. Bontcheva, and C. Scarton, "Toxic language detection in social media for Brazilian Portuguese: New dataset and multilingual analysis," arXiv preprint arXiv:2010.04543, 2020.

[27] N. Ashraf, A. Zubiaga, and A. Gelbukh, "Abusive language detection in youtube comments leveraging replies as conversational context," PeerJ Computer Science, 7, pp. 742, 2021.

[28] M. Jahan, I. Ahamed, M. R. Bishwas, and S. Shatabda, "Abusive comments detection in bangla-english code-mixed and transliterated text," In 2019 2nd International Conference on Innovation in Engineering and Technology (ICIET), pp. 1-6, IEEE, December 2019.

[29] S. Thavareesan, and S. Mahesan, "Sentiment analysis in Tamil texts: A study on machine learning techniques and feature representation," In 2019 14th Conference on industrial and information systems (ICIIS), pp. 320-325, IEEE, December 2019.

[30] S. Thavareesan, and S. Mahesan, "Word embedding-based part of speech tagging in Tamil texts," In 2020 IEEE 15th International conference on industrial and information systems (ICIIS), pp. 478-482, IEEE, November 2020.

Knowledge Graph-Based Badminton Tactics Mining and Reasoning for Badminton Player Training Pattern Analysis and Optimization

Xingli Hu[1]*, Jiangtao Li[2], Ren Cai[3]
School of Physical Education, University of Sanya, Sanya 572022, Hainan, China[1, 2]
Hainan Medical College, Haikou 571157, Hainan, China[3]

***Abstract*—As the global emphasis on sports data analysis and athlete performance optimization continues to grow, traditional badminton training methods are increasingly insufficient to meet the demands of modern high-level competitive sports. The exploration and reasoning of badminton tactics can significantly aid coaches and athletes in better comprehending game strategies, playing a vital role in the analysis and optimization of training methods. By utilizing knowledge graph-based badminton tactics mining, an approach involving heterogeneous graph splitting is employed, coupled with the incorporation of a cross-relational attention mechanism within relational graph neural networks. This mechanism assigns varying weights based on the importance of neighboring nodes across different relations, facilitating information aggregation and dissemination across multiple relationships. Furthermore, to address the challenges posed by the complexity of large-scale knowledge graphs, which feature numerous entity relationships and intricate internal structures, techniques such as training subgraph sampling, positive-negative sampling, and block-diagonal matrix decomposition are introduced. These techniques help to reduce the computational load and complexity of model training, while also enhancing the model's generalization capabilities. Finally, comparative experiments conducted on a proprietary badminton tactics dataset demonstrated the effectiveness and superiority of the proposed model improvements when reasonable parameters were applied. The case study shows that this approach holds considerable promise for the analysis and optimization of badminton players' training strategies.**

Keywords*—Badminton tactical analysis; graph neural networks; attention mechanisms; training pattern optimization; heterogeneous graph splitting; artificial intelligence

I. Introduction

As competitive sports continue to evolve, the demands on athletes to demonstrate exceptional performance in matches have become increasingly stringent. To meet these challenges, athletes need not only strong physical and psychological qualities but also advanced technical and tactical skills to identify and address their weaknesses while maximizing their strengths. Effective and scientifically sound training is crucial for enhancing athletes' competitive levels. However, the current training systems still largely rely on coaches' experience, with insufficient support from systematic data. This is particularly evident in badminton training, where the focus on technical and tactical training is relatively low, often depending on the subjective judgments of coaches and the personal feelings of athletes. This reliance on traditional methods has led to the underutilization of critical match data and a lack of scientifically based standards for evaluating technical and tactical performance.

In this context, the application of information technology is essential for improving traditional training models. However, current information technologies face significant challenges in data collection, analysis, and training optimization: data collection is often untimely, incomplete, and inaccurate, making integration difficult; data analysis lacks depth, failing to produce actionable insights; and personalized technical and tactical training programs are scarce and lack scientific backing. Additionally, the speed of information feedback does not match the fast pace of training and competition, ultimately affecting athletes' performance.

Therefore, the effective collection and utilization of training and competition data, as well as the development of personalized training programs based on this data, have become urgent tasks. With the advancement of modern database technologies and data mining techniques, these tools have increasingly been applied in the sports domain. Establishing a badminton sports database and data analysis system can significantly enhance the efficiency and depth of data collection, providing valuable insights through in-depth data analysis to inform coaches' training decisions and optimize training methods.

To fill these gaps, this paper explores how badminton tactics mining and reasoning based on knowledge graphs can assist coaches and athletes in better understanding match strategies and play a crucial role in the analysis and optimization of training methods. We propose a tactics mining approach based on heterogeneous graph decomposition, which incorporates a cross-relational attention mechanism to assign different weights in relational graph neural networks, enabling the aggregation and transmission of information across relationships. Additionally, to address the complexity challenges posed by large-scale knowledge graphs with intricate relational structures, we introduce techniques such as training subgraph sampling, positive-negative sampling, and block-diagonal matrix decomposition to reduce computational complexity and enhance model generalization. Comparative experiments conducted on a proprietary badminton tactics dataset demonstrate the effectiveness and superiority of the improvements made to our model.

*Corresponding Author

II. Literature Review

This section reviews existing related work on data mining to highlight the gaps in existing research.

A. Data Mining Techniques

Data mining technology has found widespread use in the global sports arena [1]. A well-known example is the Advanced Scout system employed by teams in the American professional basketball league, which processes game data to generate key insights, such as identifying high-percentage shooting areas and determining the most effective player rotations. This system aids head coaches in developing evidence-based training plans and making real-time strategic decisions during games [2]. In the National Football League (NFL), Schatz's study of league statistics revealed a link between league averages and the performance success of players in specific offensive roles [3]. Similarly, in the National Collegiate Athletic Association (NCAA), Jay Coleman and his team analyzed historical sports data, formulating a predictive model for game outcomes that boasts an accuracy rate of 84.5% [4].

In the realm of tennis, Damien Demaj tracked the serve trajectories of opponents and, utilizing spatial and temporal data analysis techniques, discerned patterns in player movements. This allowed athletes to gain a better understanding of their opponents' technical and tactical tendencies, ultimately improving their match performance [5][6]. Additionally, Rajiv Maheswara and his team leveraged data mining and machine learning to scrutinize the movements of players, referees, and basketballs captured by high-speed cameras during NBA games. Their analysis focused on offensive scoring efficiency and defensive effectiveness, providing coaches with critical information to structure more strategic training programs for their teams [7].

B. Traditional Knowledge Reasoning

Rule-based reasoning relies on a predefined set of rules to logically derive new knowledge, directly incorporating expert domain knowledge. While this approach is effective, it is often time-consuming to create these rules, and the system's scalability is generally limited. For instance, knowledge graphs like NELL [8] and YAGO [9] use rule-based methods to expand their knowledge bases. Meanwhile, first-order probabilistic language models focus on achieving precise "local" reasoning, proving effective in tasks such as entity resolution by extending stochastic logic programs and employing PageRank variants [10] for inference. Additionally, the TensorLog system integrates knowledge graph reasoning with gradient-based deep learning [11], achieving linear computational efficiency through a tractable reasoning process.

Ontology-based reasoning, on the other hand, generates new knowledge by interpreting, reasoning, and integrating ontologies—core structures that describe entities, attributes, relationships, and concepts. For example, a semi-automatic schema construction approach [12] addresses the complex schema challenges in RDF knowledge bases, while a system based on Markov logic networks is employed to clean and refine the original knowledge base [13]. Furthermore, the method proposed by Pujara et al. [14], which uses joint inference with probabilistic soft logic [15][16], tackles issues of noise and incomplete information in large-scale knowledge graphs. The ontology path discovery algorithm OP developed by Chen et al. further enhances knowledge graphs through optimization techniques [17].

In summary, although traditional reasoning methods can be effectively applied to knowledge graphs, ontology-based methods typically offer higher accuracy, making them suitable for scenarios where precision is critical. However, these methods face challenges with large-scale knowledge instantiation, struggling with inference efficiency and recall, while noise in the raw data can lead to inference errors, limiting the scope and applicability of these traditional approaches.

C. Distributed Knowledge Reasoning Approach

Distributed representation learning centers on transforming traditional symbolic representations into numerical representations in vector space through mapping functions as a way to mitigate dimensionality catastrophe and capture implicit connections between entities and relationships. These methods are generally divided into two main categories: rule-based reasoning and ontology-based reasoning.

The TransE model [18] is a distance-based approach that embeds entities and relations in a low-dimensional vector space to model multi-relational data. It predicts links by making the sum of the head entity vector and the relation vector as close as possible to the tail entity vector. TransE has demonstrated strong performance across several knowledge bases, surpassing advanced methods of its time, and thus gained attention. While simple and fast to train, its main limitation is that it only handles one-to-one relationships. To address this, Zhen et al. [19] introduced an improved version that considers complex relationships, such as one-to-many, many-to-one, and many-to-many. The TransH model followed, performing translation on a hyperplane to better handle diverse relationships, while maintaining similar complexity to TransE. Additionally, specific strategies were introduced to mitigate incorrect negative labels, improving performance on tasks like link prediction and fact extraction in knowledge graphs like WordNet and Freebase.

The RESCAL model, proposed by Nickel et al. [20], is a tensor decomposition-based inference method that enables collective learning through its latent components and offers efficient factorization. RESCAL demonstrated significant improvements in speed and accuracy over other tensor methods. Based on RESCAL, Chang et al. [21] proposed TRESCAL, which excels in relationship extraction by leveraging entity type information, allowing faster and more accurate discovery of new relations in databases.

Inference methods based on distributed representations are more advanced than traditional methods. The TransE series, known for its simplicity and effectiveness, has become a research focus, achieving significant results despite limited further research space. The RESCAL series, favored for its interpretability and performance, holds great potential, despite higher model complexity.

D. Research Gaps

Although progress has been made in applying graph neural networks and association rule mining to knowledge graphs [22], significant gaps remain in their application to badminton tactical

reasoning and training. Current algorithms fail to fully consider dynamic player positioning, opponent adaptability, and the data requirements of input models.

1) Model adaptability and flexibility: Existing graph neural networks can simulate some aspects of badminton tactics but struggle with the complexity and real-time nature of tactical changes. These models require deeper integration of tactical theory and practical experience, and current approaches are often inefficient due to heavy reliance on data annotation and feature design.

2) Dataset standardization: The lack of unified standards for data collection and preprocessing in badminton tactics leads to inconsistencies and hinders the generalizability of different studies. Without standardized data representation, tactical reasoning methods face challenges in coherence and adaptability.

3) Complex tactical reasoning: Current techniques struggle with recognizing and reasoning about complex tactical combinations and rapidly changing scenarios. They often underperform in long-term tactical evolution and opponent strategy adaptation, requiring further optimization.

In summary, future research should focus on developing more adaptable models and standardized datasets to enhance tactical reasoning, thereby improving technical and tactical teaching and game strategy analysis in badminton.

III. Ideas for Improving Reasoning in Very Large Scale Knowledge Graphs

A. Heterogeneous Graph Splitting

Entities and relationships in a knowledge graph often exhibit diverse types and properties, creating a heterogeneous graph structure. This heterogeneity is crucial in reasoning since different types of entities and relationships may involve distinct rules, constraints, and semantics.

In this study, we employ a knowledge graph splitting approach. Heterogeneous graph splitting involves dividing a large, complex graph into smaller, more manageable subgraphs based on specific rules. This method enhances the scalability, maintainability, and processing efficiency of large knowledge graphs, particularly in distributed environments where managing smaller subgraphs is more practical.

The process of splitting a heterogeneous knowledge graph involves several key steps:

- Defining Splitting Rules: Based on the characteristics and application needs of the knowledge graph, we establish rules for splitting—such as dividing by entity types, relationship types, or attributes.
- Constructing Topologies: Using the defined rules, we construct the topology for each subgraph. This can be achieved through graph-based clustering algorithms or similarity calculations based on meta-paths.
- Performing the Split: Entities and relationships in the original graph are divided into multiple subgraphs according to the constructed topologies.
- Storing and Managing Subgraphs: Each subgraph is stored in different computing nodes or distributed storage systems, with appropriate management and maintenance.

In our study, we focus on splitting subgraphs based on relationship types. During the reasoning process, these subgraphs are embedded to better capture the distinct semantic information inherent in the heterogeneous graph structure. For example, as illustrated in Fig. 1, an original directed graph with five nodes and two types of relationships can be split into two subgraphs, each retaining the directed nature of the original structure. This allows for more precise reasoning based on the specific types of relationships in each subgraph.

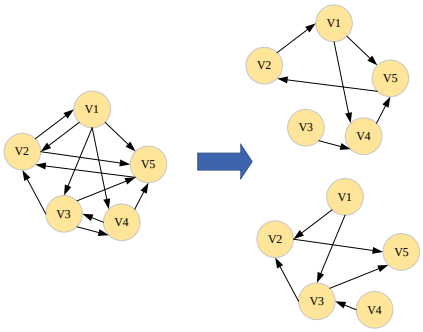


Fig. 1. Schematic diagram of knowledge graph heterogeneous graph splitting.

B. Training Subgraph Sampling

Training graph neural networks on large-scale datasets presents challenges, such as high computational complexity and issues like gradient vanishing or explosion. To address these, this study employs random sampling to generate subgraphs for training. Random sampling reduces noise by excluding irrelevant nodes and edges, thereby lowering computational and storage demands and improving training efficiency. However, tuning sampling parameters is necessary to balance subgraph size with the importance of sampled components.

Additionally, positive and negative sampling is used to preprocess training data by selecting real (positive) and non-existent (negative) edges or nodes. This approach offers several benefits:

- Addressing Data Imbalance: By balancing the number of positive and negative examples, the model avoids bias towards predicting positive instances, enhancing its overall prediction capabilities.
- Reducing Computation: Sampling limits the data used for training, cutting down on computational load and improving efficiency.
- Improving Generalization: Positive and negative sampling enables the model to learn a broader range of features, enhancing its adaptability to different graph datasets.

In summary, the introduction of positive and negative sampling in this study is crucial for addressing data imbalance, boosting computational efficiency, and enhancing the model's generalization ability.

IV. Improved Knowledge Graph Link Prediction Algorithm

In a knowledge graph, link prediction involves forecasting the relationship between two entities based on the existing graph structure, often referred to as relationship extraction. This process is crucial for expanding the knowledge graph, as manually labeling every entity and relationship is impractical given their vast number. Link prediction enables the automatic discovery of new entities and relationships from large volumes of unlabeled data, thereby enhancing the graph's scale and richness. The primary objective is to predict potential relationships between entities or to identify links between entities and relationships. This task has broad applications in fields such as information retrieval, natural language processing, recommender systems, and Q&A, where it helps to uncover correlations within the knowledge graph, thereby improving the effectiveness of these systems.

In this study, building on the concepts outlined in Fig. 1, we aim to enhance the knowledge graph link prediction algorithm by integrating key techniques such as knowledge graph embedding, data preprocessing, and graph neural networks. This approach will enable convolutional learning of entities and relationships, utilize the DistMult decoder for scoring, and apply association rule mining to select the ternary elements for evaluation. The overall architecture of the improved algorithm is depicted in Fig. 2.

A. Knowledge Graph Input to the Algorithm

Before modeling a knowledge graph, the entities and relationships within it must be converted into vector representations. This conversion is crucial because it allows us to use vector space distance and similarity measures to assess relationships between entities and infer connections between unknown entities—this process is known as knowledge graph embedding.

As illustrated in Fig. 3, the entity set in the knowledge graph is first encoded as text. These encoded entities are then mapped to corresponding sequence numbers, which typically start from 0 and increment sequentially. A dictionary is constructed to link each sequence number to an entity in the knowledge graph. During model inference, these sequence numbers are consistently used to process entity information. Additionally, a dictionary is created to map sequence numbers to vector representations, ensuring that each sequence number corresponds to a low-dimensional vector representing the final entity.

For the edge set in the knowledge graph, each edge type e_i is transformed into a relation r_i. Similar to entities, a dictionary is created to map relations to sequence numbers, and another dictionary links these sequence numbers to relation vectors. Through this process, the knowledge graph is transformed into $G = (V, \varepsilon, R)$, converting the knowledge within the graph into vector representations that can be used as inputs for subsequent algorithms.

B. Node Information Aggregation based on Node Cross-Relational Attention Mechanism

After transforming the entities and relationships in the knowledge graph into vectors, the structure of graph neural network is used to aggregate the information of the nodes to realize the task of link prediction in the existing knowledge graph. The entire knowledge graph is represented as a set of $G = (V, \varepsilon, R)$ vectors, where $\{V\} = \{1,2,3,\ldots,n\}$ is the set of all the entities in the knowledge graph, $\{\varepsilon\}$ is the set of all the edges, and $\{R\}$ is the set of relationships. Assuming that the head node of the triad is $v_i \in V$, the tail node is $v_j \in V$, and the relation between nodes is $r \in R$, then this triad can be represented as $(v_i, r, v_j) \in \varepsilon$.

In GCN, the embedding vector of a node can be computed by using the Eq. (1).

$$h^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}h^{(l)}W^{(l)}\right) \quad (1)$$

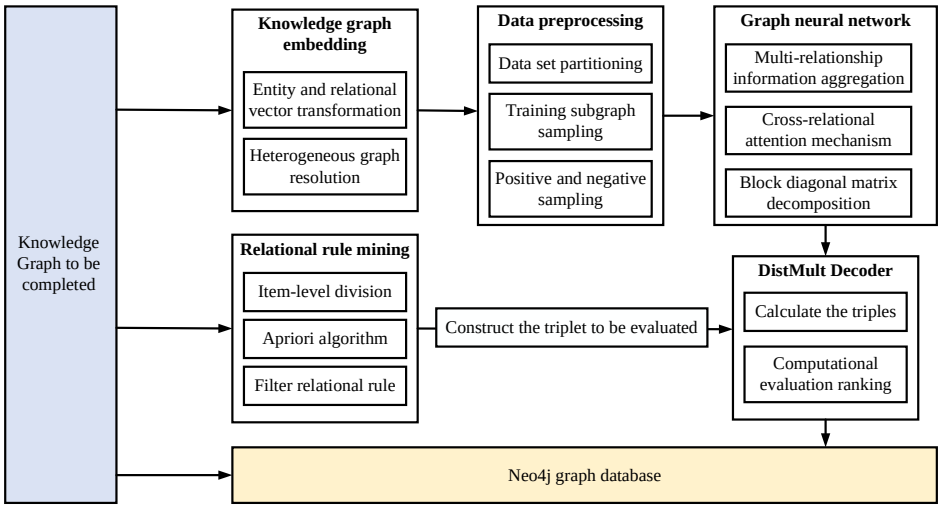


Fig. 2. Algorithm architecture diagram.

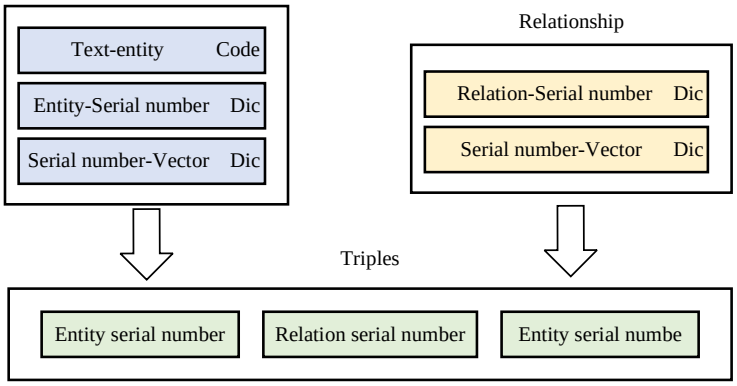


Fig. 3. Schematic of the conversion of knowledge graph triples to vectors.

where, $h^{(l)}$ is the node embedding vector of layer l, $W^{(l)}$ is the trainable weight matrix of layer l, $\tilde{A} = A + l$is the sum of the adjacency and self-connectivity matrices of the graph, $\widetilde{D}$ is the degree matrix of $\tilde{A}$, and σ is the activation function.

Based on Eq. (1), the process of GCN information aggregation can be expressed by Eq. (2).

$$h_i^{(l+1)} = \sigma\left(\sum_{j\in N_i} \frac{1}{\sqrt{d_i}d_j} h_j^{(l)} W^{(l)}\right) \quad (2)$$

where, $h_i^{(l+1)}$ denotes the embedding vector of node i at layer $l+1$, $W^{(l)}$is the trainable weight matrix, N_i is the set of neighbors of node i, $\widetilde{d}_i$ and $\widetilde{d}_j$ are the degrees of node i and node j in matrix $\tilde{A}$, respectively.

As can be seen from Eq. (2), in GCN, the information of nodes converges in different relations sharing the same weight matrix W, so GCN actually treats the graph as a homomorphic graph for processing, in this study, we will consider the heteromorphism in the knowledge graph, and we have already transformed the entities and relations in the knowledge graph into vectors in the above paper and constructed the mapping between the serial numbers and vectors of the entities and relations, using the correspondence between serial number and vector, which makes it easier to split the knowledge graph into different heteromorphic graphs for processing according to different relations, which is reflected in the fact that different relations construct different maps between entities and relations. Using the correspondence between the serial number and the vector, we can easily split the knowledge graph into different heterogeneous graphs according to different relationships, which is reflected in the construction of different weight matrices for different relationships, instead of using the same weight matrix for all relationship types as in the case of GCN.

On the basis of GCN, considering the heterogeneity of the graph, the information is aggregated for different types of edges between nodes, as shown in Eq. (3).

$$h_i^{(l+1)} = \sigma\left(\sum_{r\in R}\sum_{j\in N_i^N} \frac{1}{c_{i,r}} W_r^{(l)} h_j^{(l)} + W_0^{(l)} h_i^{(l)}\right) \quad (3)$$

Where, $h_i^{(l+1)}$ denotes the embedding of node i in the $l+1$ layer, σ is the activation function, R is the set of relations, $N_i^r$is the set of neighboring nodes of node i under relation r, $c_{i,r}$is the normalization constant, and $W_r^{(l)}$ is the weight matrix of relation r. Since the information of nodes themselves needs to be taken into account, the self-loop of nodes is introduced. The self-loop of a node can be treated as a special type of edge, and the self-loop weight matrix is denoted by $W_0^{(l)}$.

As can be seen from Eq. (3), for each relationship, a separate weight matrix W_r is computed. In large-scale knowledge graphs, due to the large number of relationships, the number of parameters in the inference process is also unusually large, which further increases the complexity of the inference process. Specifically, assuming that a graph G contains N nodes and M relationship types, each relationship type has a corresponding weight matrix W_r, size $D_l \times D_{l-1}$, where D_l is the feature dimension of the current layer of the relationship type, and D_{l-1} is the feature dimension of the previous layer. Since each node has different embedding vectors under different relation types, these weight matrices need to be computed separately at each node, resulting in an extremely large number of parameters.

In order to reduce the complexity of the model and improve the generalization performance, the diagonal decomposition of the weight matrix is performed, which transforms the complex weight matrix into a block diagonal matrix, thus reducing the number of parameters and improving the generalization performance of the model. Specifically, as shown in Eq. (4), where the size of $Q_{br}^{(l)}$ matrix is $\left(d^{(l+1)}/B\right) \times \left(d^{(l)}/B\right)$, the block diagonal decomposition of the weight matrix through a series of matrices of the summation, which greatly reduces the size of the parameters of the weight matrix and simplifies the computation process.

$$W_r^{(l)} = \bigoplus_{b=1}^{B} Q_{br}^{(l)} = diag(Q_{1r}^{(l)}, \ldots, Q_{Br}^{(l)}) \quad (4)$$

In the above node information aggregation process, each relation type uses an independent weight matrix for information transfer, and the node embedding is obtained by splicing the information of each relation type. This approach may lead to the uneven contribution of each relationship type in node embedding, and the contribution of some relationship types may be masked or weakened. To solve this problem, this study introduces the Node-level Across Relation Attention mechanism, which adjusts the weight of each relation type in the node

embedding by learning a node-level attention vector for each node.

Before calculating the node's attention mechanism, firstly, for the similarity between node i and node j under a specific relation r, the weight matrix corresponding to relation r is multiplied and spliced with node i and node j, respectively, and then multiplied with a trainable attention vector, and then the similarity between node i and node j under relation r is finally obtained by an activation function. As shown in Eq. (5), where e_{ij}^r denotes the similarity between node i and node j under relation r, $\vec{a}^r$ is a trainable attention vector, $\vec{W}_r$ is the weight matrix under relation r, and $\oplus$ denotes the vector splicing operation.

$$e_{ij}^r = \mathrm{ReLU}\left(\vec{a}^r \cdot \left[\vec{W}_r h_i^{(1)} \oplus \vec{W}_r h_j^{(1)}\right]\right) \quad (5)$$

The calculation process of attention is shown in Fig. 4. The final information aggregation process of a single node is shown in Fig. 5.

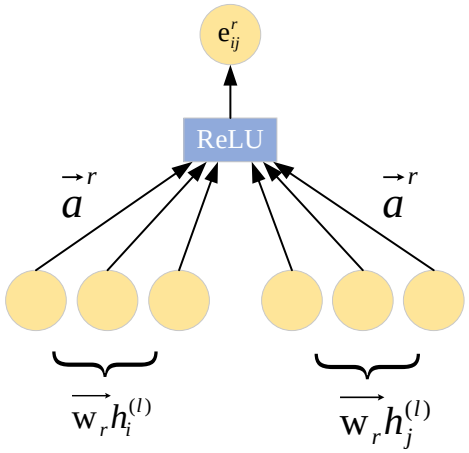


Fig. 4. Attention calculation process.

Based on Eq. (5), in the iterative process of graph neural network, each node in the knowledge graph calculates the similarity of its neighboring nodes under all the relationships, and then uses the similarity to calculate the attention of the node's neighboring nodes, and ultimately the information convergence of nodes in each layer is shown in Eq. (6). The whole information aggregation process is shown in Fig. 6, and finally the whole network outputs the embedding vector representation of nodes and relationships.

$$h_i^{(1+1)} = \sigma\left(\sum_{r\in R}\sum_{j\in N_i^r}\frac{exp\left(\mathrm{ReLU}\left(\vec{a}^r\cdot\left[\vec{W}_r h_i^{(l)}\oplus\vec{W}_r h_j^{(1)}\right]\right)\right)}{\sum_{k\in N_i^r} exp\left(\mathrm{ReLU}\left(a^r\cdot\left[\vec{W}_r^h h_i^{(l)}\oplus\vec{W}_r h_k^{(1)}\right]\right)\right)} W_r^{(l)} h_j^{(l)} + W_0^{(l)} h_i^{(l)}\right) \quad (6)$$

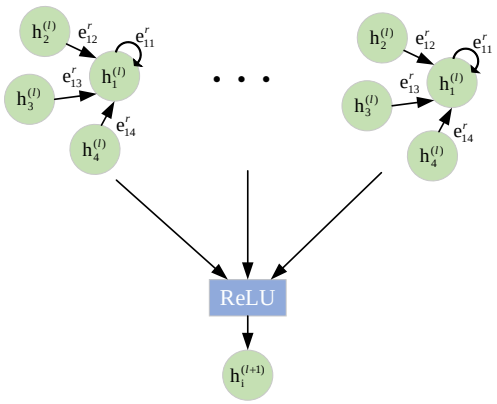


Fig. 5. Individual node final information aggregation process.

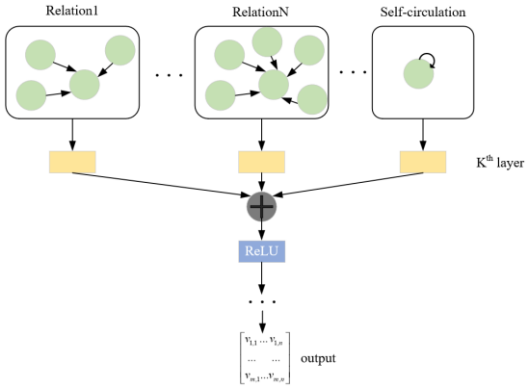


Fig. 6. Schematic diagram of node information aggregation process.

C. Training Sampling and Loss Functions

In the link prediction task, dealing with an entire knowledge graph can be challenging due to its size, often exceeding the capacity of a single GPU's memory. To address this, we sample smaller subgraphs for training. By using random sampling, we can maintain the integrity of the graph's structure while avoiding memory bottlenecks associated with processing large-scale knowledge graphs. This study employs uniform random sampling, where each node or edge is selected with equal probability. This method is straightforward, offering good randomness and repeatability, and can be executed using a random number generator.

The sampling process involves the following steps:

1) Determine the sample size: Set a hyperparameter, batch_size, to define the number of nodes or edges to sample.

2) Initialize the sample set: Create a set to store the sampled nodes or edges.

3) Random sampling: For each node or edge, generate a random number; if it falls below the sampling probability, include the node or edge in the sample set.

4) Return the sample set: Use this set for subsequent training or other tasks.

After random sampling, a training subgraph is obtained. To further enhance prediction accuracy and generalization, the study introduces positive and negative sampling. For each node pair (i, j), both positive samples (existing links) and negative samples (non-existent links) are generated. Negative samples are constructed using a random sampling strategy, with a negative sampling factor μ, where the ratio of positive to negative samples is $1:\mu$.

The loss function for the final model training is shown in Eq. (7) as follows.

$$L = -\sum_{(i,j)\in E} log\,\sigma\,(y_{(i,j)}) - \sum_{(i,j)\in \tilde{E}} log\,\sigma\,(\tilde{y}_{(i,j)}) \quad (7)$$

Where, E denotes the set of positive sample edges, E denotes the set of negative sample edges, and $y_{(i,j)}$ denotes the predicted probability of the existence of an edge between node *i* and node *j*, which can be computed by Eq. (8).

$$y_{(i,j)} = \frac{1}{k}\sum_{}^{t=1}{}^{k} \sigma(h_i^{(L)} \cdot r_i^{(L)} h_j^{(L)}) \quad (8)$$

where *k* denotes the number of negative samples sampled, $h_i^{(l)}$ and $h_j^{(l)}$ denote the embedding vectors of node *i* and node *j* in the last layer, and $r_t^{(L)}$ denotes the weight vector of relation type r_t in the Lth layer. σ denotes the sigmoid function, which is used to map the predicted values to a range between 0 and 1. The first term of the loss function is the cross-entropy loss for positive samples, indicating that the higher the probability that the model predicts the existence of positive edges, the lower the loss. The second term is the cross-entropy loss of negative samples, which means that the higher the probability that the model predicts that the negative sample edge does not exist, the lower the loss. By minimizing the loss function, it allows the model to predict the likelihood of edge existence more accurately between nodes.

V. Comparative Experiments and Analysis

This section presents comparative experiments to evaluate the model's effectiveness, focusing on the experimental environment, dataset, evaluation criteria, control group design, and final result analysis to validate the improvements made in this study.

A. Experimental Environment and Datasets

Due to the large knowledge graph dataset used in this experiment and the complexity of model computation, the experiment needs to be carried out on GPUs. The server used for the experiment is the one equipped in the lab, Windows 10 64bit system, 16GB RAM and equipped with high-computing-power GPUs. In terms of software, Python was used as the programming language for the model, PyCharm was used to compile the software, and PyTorch was used as the basic implementation library for the model.

This paper presents a custom-built knowledge graph dataset named BadmintonKG, specifically created for the badminton domain. It is designed to support applications such as tactical reasoning, match analysis, and training assistance. BadmintonKG includes 9,742 entities, 135 types of relationships, and 198,563 triples (i.e., relationships between entities). Unlike other common knowledge graph datasets, BadmintonKG focuses on the specific domain of badminton, covering various entities and relationships such as players, tactics, courts, techniques, and coaching styles. The training, validation and test sets are divided as shown in Table I.

TABLE I. BadmintonKG Data Set Delineation Table

Dataset segmentation	Entities	Relations	triples
Original dataset	9,742	135	198,563
Training set	9,742	135	176,421
Validation set	9,742	135	19,989
Test set	9,742	135	11,072

B. Evaluation Criteria and Control Group Design

MRR (Mean Reciprocal Rank) is one of the commonly used evaluation metrics in the task of knowledge graph link prediction. The core idea of MRR is to find the true tail entity *t* among all possible tail entities *t`* for each test ternary (h,r,t), and compute the inverse of its score ranking. Finally, the average of the ranked inverts of all the test triples is used as the MRR score of the model. This is calculated as shown in Eq. (9).

$$MRR = \frac{1}{|T|}\sum_{(h,r,r)\in T}\frac{1}{\text{rank}_{(hr,t)}} \quad (9)$$

In knowledge graph link prediction, Hits@ N is a commonly used evaluation metric to measure whether the algorithm can correctly predict the correct entity or relation in the test set among the first N candidate entities or relations. Specifically, assuming that the correct answer for each ternary (h,r,t) in the test set is *t*, then for each (h,r) pair, we can sort all its possible entities according to the algorithm's prediction scores in descending order and compute whether the first *N* predicted entities contain the correct answer *t*. If the correct answer is *t*, then it is called a hit. If it does, it is called a hit (hit), otherwise it is called a miss. The final hit rate is the average hit rate of all test triples among the first N candidate entities. This is shown in Eq. (10).

$$\text{Hits@}N = \frac{1}{|T|}\sum_{(h,r,t)\in T}\Pi(\text{rank}_{(h,r,t)} \le N) \quad (10)$$

To validate the effectiveness of the proposed knowledge graph inference algorithm based on graph neural networks and association rule mining, experiments will be conducted on BadmintonKG. The experimental model will be compared with three other neural network models: Graph Convolutional Network (GCN), Graph Attention Network (GAT), and Relational Graph Convolutional Network (R-GCN). The comparison will focus on the evaluation metrics MRR and Hits@N, and the impact of different training subgraph sizes on model performance will also be assessed.

Knowledge Graph Link Prediction with GCN: GCN is a graph convolutional neural network that treats the knowledge graph as a homogeneous graph, using node neighbors to perform convolution. It aggregates neighbor information with a weighted adjacency matrix, enabling it to learn node representations through multi-layer convolution. GCNs are particularly effective for graphs with similar node features.

Knowledge Graph Link Prediction with GAT: GAT extends GCN by incorporating an attention mechanism that weights the importance of each neighboring node. This allows for the aggregation of neighbor information with varying degrees, resulting in richer and more accurate node representations.

Knowledge Graph Link Prediction with R-GCN: R-GCN is designed for multi-relational graphs, using relational matrices in its convolution operations. Unlike traditional GCNs, R-GCN employs a learnable convolution kernel that adjusts to different types of relations, making it more effective for handling multi-relational data.

C. Experimental Results and Analysis

In this control group, this experimental model is compared with the other two models on two datasets, and the number of model iterations for the experiments is set to 6000, the stochastic inactivation rate is set to 0.2, the learning rate is set to 0.01, the output dimension of the hidden layer is set to 500, and the sampling mode of the training subgraphs is set to uniform. The test set evaluation results from the above training results are plotted in a table, as shown in Table II.

The above table was plotted as a line graph as shown in Fig. 7 and 8. From the above experimental data performance, the model in this study shows excellent performance on both datasets, and outperforms the other three models in all indicators.

In the BadmintonKG dataset, when the size of the training subgraph is 30000, this model outperforms the GCN, GAT, and R-GCN models in all the metrics, but the difference is relatively small, and when the size of the training subgraph is 80000, the MRR metrics of this model reaches 0.2753, which is higher than that of 0.1910 for the GCN, GAT, and R-GCN models, 0.2503 and 0.2653, which is about 40% higher than that of the GCN model, 10% higher than that of the GAT model, and 3.8% higher than that of the R-GCN model.

TABLE II. SUMMARY OF EXPERIMENTAL RESULTS

BadmintonKG	The size of the training subgraph is 30000				The size of the training subgraph is 80,000			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
Our	0.2458	0.1576	0.2677	0.4249	0.2753	0.1811	0.3066	0.4638
GCN	0.1673	0.0873	0.1927	0.3201	0.1910	0.0990	0.2283	0.3684
GAT	0.2144	0.1283	0.2373	0.3855	0.2503	0.1610	0.2695	0.4403
R-GCN	0.2412	0.1404	0.2531	0.4085	0.2653	0.1715	0.2856	0.4536

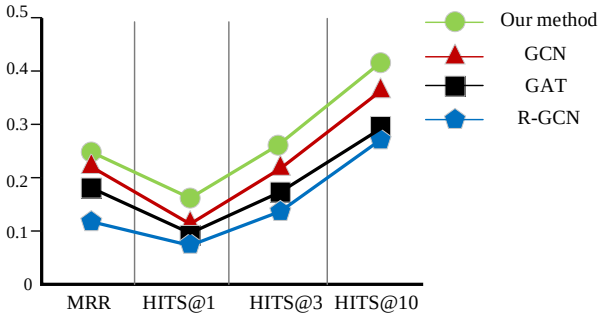


Fig. 7. BadmintonKG comparison of models (training subgraph size 30000).

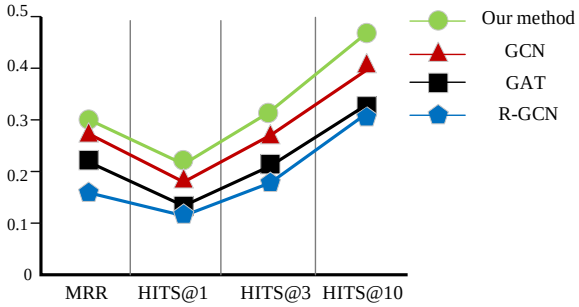


Fig. 8. BadmintonKG comparison of models (training subgraph size 80000).

Overall, the model in this study achieves better performance than the link prediction models based on GCN, GAT and R-GCN in the knowledge graph link prediction task, and shows good robustness and generalization ability, and the experimental results in this study prove the effectiveness and superiority of the proposed model in the knowledge graph link prediction task.

VI. CONCLUSION

In this study, we applied knowledge graph and graph neural network techniques to the mining and reasoning of badminton tactics, demonstrating the effectiveness of this approach in analyzing and optimizing players' training patterns. Technically, the introduction of training subgraph sampling, positive and negative sampling, and block diagonal matrix decomposition significantly reduced the computational complexity, improved the handling of large-scale knowledge graphs, and enhanced the model's generalization capabilities. These techniques not only optimized the data processing workflow but also improved the model's performance and accuracy in practical applications.

Through comparative experiments on a proprietary badminton tactics dataset, we validated the superiority of the proposed method over traditional approaches in tactical reasoning and training optimization. The results showed that our model more accurately predicted and reasoned about tactical changes in matches, offering strong scientific support for badminton training and competition. In conclusion, this study provides a novel technical approach for tactical analysis in badminton and offers a potential methodological reference for other sports.

ACKNOWLEDGMENT

Thanks to all the people and organizations that contributed to this study.

REFERENCES

[1] Fu T, Li P, Liu S. An imbalanced small sample slab defect recognition method based on image generation[J]. Journal of Manufacturing Processes, 2024, 118: 376-388.

[2] Fu T, Liu S, Li P. Intelligent smelting process, management system: Efficient and intelligent management strategy by incorporating large language model[J]. Frontiers of Engineering Management, 2024: 1-17.

[3] Schatz, A. Pro Football Prospectus: Statistics, Analysis, and Insight for the Information Age[M]. 2006 Edition. Workman Publishing Company,2006.

[4] Coleman, J.&A. Lynch. NCAA Men's Basketball Tournament Score Card[EB/OL].

[5] Fu T, Liu S, Li P. Digital twin-driven smelting process management method for converter steelmaking[J]. Journal of Intelligent Manufacturing, 2024: 1-17.

[6] Damien Demaj.Using spatial analytics to study spatio-temporal patterns in sport[EB/OL]. http://blogs.esri.com/esri/arcgis/2013/02/19/using-spatial-analytics-to-study-spatio-temporal -patterns-in-sport.

[7] Maheswaran R,Chang Y H,Henehan A,et al.Deconstructing the rebound with optical tracking data[C].The MIT Sloan Sports Analytics Conference.Boston,2012.

[8] Carlson A, Betteridge J, Kisiel B, et al. Proceedings of the Twenty-Fourth AAA I Conference on Artificial Intelligence (AAAI-10) Toward an Architecture for NeverEnding Language Learning[JI]. 2011.

[9] Suda M, Weidenbach C, Wischnewski P . On the Saturation of YAGO[C]// Inter national Conference on Automated Reasoning. Springer-Verlag, 2010.

[10] Page L, Brin S , Motwani R, et al. The PageRank citation ranking: Bringing ord er to the Web. Technical report. 1919.

[11] Cohen W W . TensorLog: A Differentiable Deductive Database:, 10.48550/arXiv. 1 605.06523[P].2016.

[12] L Bühmann, Lehmann J . Pattern Based Knowledge Base Enrichment[C]// 12th I nternational Semantic Web Conference, 21-25 October 2013, Sydney, Australia. Springer, Berlin, Heidelberg, 2013.

[13] Jiang S , Lowd D , Dou D . Learning to Refine an Automatically Extracted Kno wledge Base Using Markov Logic[C]// IEEE International Conference on Data Minin g;ICDM2012. 0 .

[14] Pujara J , Miao H , Getoor L, et al. Knowledge Graph Identification[C]// Intern ational Semantic Web Conference. Springer-Verlag New York, Inc. 2013.

[15] Zheng H, Liu S, Zhang H, et al. Visual-triggered contextual guidance for lithium battery disassembly: a multi-modal event knowledge graph approach[J]. Journal of Engineering Design, 2024: 1-26.

[16] Kimmig A, Bach S , Broecheler M , et al. A Short Introduction to Probabilistic Soft Logic[J]. Dec-2012, 2013.

[17] Chen Y , Goldberg S , Wang D Z, et al. Ontological Pathfinding[C]// the 2016 I nternational Conference. ACM, 2016.

[18] Bordes A, Usunier N , Garcia-Duran A, et al. Translating Embeddings for Mod eling Multi-relational Data[C]// Neural Information Processing Systems. Curran Asso ciates Inc. 2013.

[19] Zhen W , Zhang J , Feng J , et al. Knowledge Graph Embedding by Translating on Hyperplanes[C]// National Conference on Artificial Intelligence. AAAI Press, 201 4.

[20] Nickel M , Tresp V , Kriegel H P. A Three-Way Model for Collective Learning on Multi-Relational Data[C]// International Conference on International Conference o n Machine Learning. Omnipress, 2011.

[21] Chang K W , Yih W T , Yang B, et al. Typed Tensor Decomposition of Knowle dge Bases for Relation Extraction[J]. 2014.

[22] Liu S, Zheng P, Xia L, et al. A dynamic updating method of digital twin knowledge model based on fused memorizing-forgetting model[J]. Advanced Engineering Informatics, 2023, 57: 102115.

Analysis and Usability Evaluation of Virtual Reality in Cultural Landscape Promotion Platform Application

A Case Study from Dongao Ding Village, Dongtou District

Yufang Huang[1], Yin Luo[2], Jingyu Zheng[3]*, Xunxiang Li[4]
College of Fine Arts and Design, WenZhou University, WenZhou 325035, ZheJiang, China[1, 3, 4]
WenZhou University, WenZhou 325035, ZheJiang, China[2]

***Abstract*—In order to improve the efficiency of analysing the application of virtual reality technology in cultural landscape promotion platform and enhance the accuracy of usability assessment, a feasibility assessment method based on MA-BiGRU for virtual reality technology cultural landscape promotion platform is proposed. Firstly, it analyses the application of virtual reality technology in cultural landscape promotion platform and designs the application feasibility assessment; secondly, it combines MA algorithm and BiGRU network, and proposes a usability assessment algorithm of virtual reality technology based on MA-BiGRU model; lastly, it analyses the feasibility and validity of the proposed method by using actual cases. The results show that, compared with other models conducted, the proposed method has a higher assessment and prediction accuracy, and it also effectively assists in completing the virtual reproduction of the cultural landscape promotion platform of Dongao Deng Village, Dongtou District, Wenzhou City, and improves the design effect of virtual reality technology in the cultural landscape promotion platform.**

Keywords*—Virtual reality technology; cultural landscape promotion platform; usability assessment algorithm; Mayfly algorithm

I. INTRODUCTION

With the continuous breakthrough of human science and technology, the visual experience of two-dimensional plane has slowly failed to meet the needs of people's life, so the three-dimensional virtual technology has gradually entered people's lives [1]. Virtual reality technology as one of the most promising technologies, with the development and improvement of hardware and software technology at this stage, has been widely used in a variety of industries, but virtual reality technology in the field of cultural landscape is lacking [2]. Currently, the main status quo of cultural landscape promotion research is reflected in the rigidity of cultural communication form, the lack of commercial profit model, combined with virtual reality technology to study the cultural landscape promotion platform not only improves the utilisation rate of cultural landscape resources, but also accelerates the speed of cultural landscape dissemination, but also realises the inheritance, innovation and development of cultural landscape [3]. The application of virtual reality technology in the cultural landscape promotion platform is mainly divided into three aspects: the application of virtual reality technology, the analysis of the application of virtual reality technology and the assessment of the usability of virtual reality technology [4].

*Corresponding Author

The application of virtual reality technology in cultural landscape promotion platform is mainly the combination of virtual reality technology and cultural landscape promotion platform design; virtual reality technology application analysis is mainly to analyse the application of virtual reality technology in cultural landscape promotion platform index mention analysis; virtual reality technology usability assessment refers to the construction of virtual reality technology in cultural landscape promotion platform application analysis and assessment model, prediction and assessment of application situation. Cai and Jin [5] studied the combination of virtual reality technology and the protection and inheritance of village and town cultural landscape, and put forward the application strategy. Li and Wang [6] studied the application of virtual reality technology in cultural theme parks, and combined the elements of jin yong's martial arts world. Mo [7] studied the method of reproducing historical and cultural landscapes based on virtual reality technology. Yang and Sun [8] addressed the problem of virtual reality display in fuzimiao, and researched the historical humanistic landscape of cultural communication and tourism business promotion strategy. Dimitar et al [9] studied the usability assessment method of virtual reality technology based on machine learning, and at the same time for the machine learning algorithms fall into the local optimal problem using intelligent optimisation algorithms to find the optimal. By analysing the above literature, although there are some research results on the combination of virtual reality technology and cultural landscape promotion platform, there are some problems: 1) the role analysis of virtual reality technology on the design of cultural landscape promotion platform is not clear enough; 2) the extraction of indicators for the analysis of the application of virtual reality technology is not comprehensive and systematic enough; 3) the accuracy of the usability assessment of virtual reality technology is low; 4) the application of virtual reality technology fails to achieve the design efficiency [10].

In view of the above problems, this paper proposes a usability assessment method for cultural landscape promotion platform of virtual reality technology based on mayfly algorithm optimised two-way gated recurrent unit neural network. Focusing on the application of virtual reality technology in the design of cultural landscape promotion platform, the application process of VR in cultural landscape promotion platform is analysed, and the relevant usability assessment indexes are extracted; for the usability assessment of the cultural landscape promotion platform based on virtual reality technology,

combining with the mayfly algorithm [11] to optimize the BiGRU neural network parameters [12], a MA-BiGRU-based feasibility assessment model of cultural landscape promotion platform based on virtual reality technology; through case study, the model proposed in this paper improves the feasibility assessment accuracy of cultural landscape promotion platform and verifies the feasibility of virtual reality technology in cultural landscape promotion platform.

This paper begins with a review of the status quo of cultural landscape promotion and how VR can address existing gaps. It then elaborates on the integration of MA and BiGRU for the usability assessment model, followed by an analysis of VR technology's impact on the design of cultural landscapes. The study incorporates actual case analysis from Dongao Ding Village, demonstrating the improved assessment and prediction accuracy of the proposed method compared to traditional models.

The paper is structured as follows:

- An overview of VR applications in cultural landscape promotion platforms.
- Explanation of the Mayfly Algorithm (MA) and the BiGRU model.
- The construction of a usability assessment method integrating MA and BiGRU.
- Application and evaluation of the proposed model using the Dongao Ding Village case.
- A discussion of results and performance improvements in the proposed approach.

II. Virtual Reality Application in Landscape Promotion

A. Virtual Reality

The use of Virtual Reality (VR) technology in cultural landscape promotion platforms is becoming more widespread, enabling users to experience cultural and tourism content in an immersive way through simulated environments.

1) VR technology: VR virtual reality technology can be regarded as a major development direction of simulation technology, which actually contains a variety of different technologies, such as human-computer interface, multimedia sensing, networking, etc. [13], as shown in Fig. 1. And the performance mode of VR virtual reality technology also contains several parts, such as environment simulation, external perception, natural skills, and sensing devices.

2) VR technology applications: According to the principle of VR technology, the application of VR technology in the cultural landscape promotion platform is mainly manifested in the following aspects [14]: 1) the construction of intelligent scenic spots; 2) the promotion of cultural and tourism virtual reality applications; 3) the digital promotion of national cultural industry parks; 4) the construction of digitally empowered cultural and tourism scenarios; and 5) the immersive tour experience of VR technology, as shown in Fig. 2.

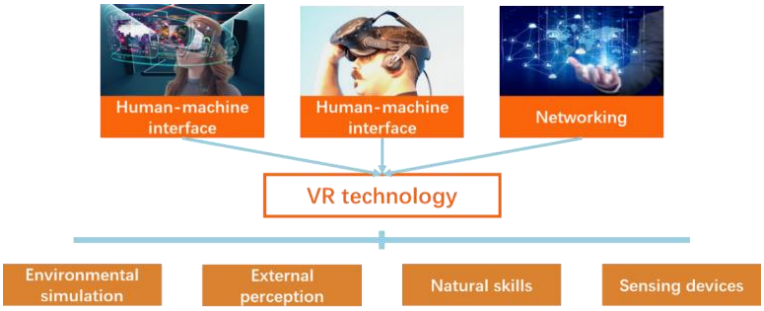


Fig. 1. Principle of VR technology.

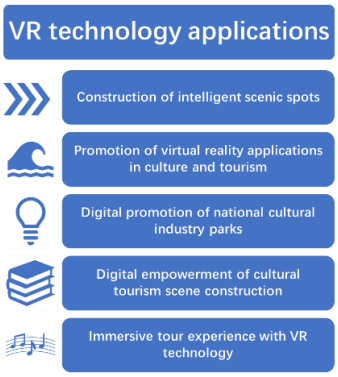


Fig. 2. Application of VR technology in the cultural landscape promotion platform.

3) Steps for the application of VR technology in cultural landscape promotion platforms: Virtual reality technology can not only bring visitors a rich and shocking visual impact, but also need to use virtual reality technology to clearly convey the information of the cultural landscape to the visitors. This paper takes the virtual reproduction of cultural landscape as a case study, based on the reality, adopts the multi-user interactive 3D modelling method of GIS, and applies the VR technology to the problem of cultural landscape reproduction [15], the specific steps are shown in Fig. 3.

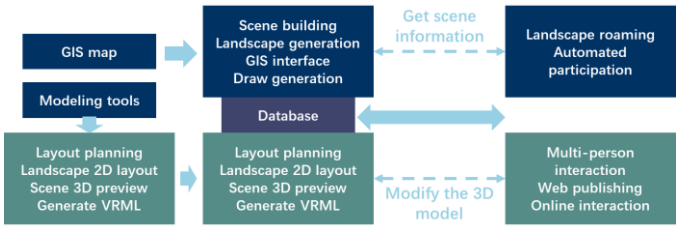


Fig. 3. Steps in reproducing the landscape virtual environment.

B. Analysis of VR Application in Cultural Landscape Promotion Platforms

According to the principles of demand, scientific, systematic and quantifiable (Fig. 4), by analysing the steps of applying VR technology in the cultural landscape promotion platform, this paper extracts the application analysis indicators from the four aspects of landscape environment, image, behaviour and spirit, as shown in Fig. 5. (1) Landscape environment includes indicators such as topography and geomorphology, climate, water bodies, flora and fauna, etc.; (2) Landscape image includes indicators such as residents, squares, historical sites, sculpture vignettes, etc.; (3) Landscape behaviour includes indicators such as daily behavioural habits, territorial construction methods, and landscape industry ideas, etc.; and (4) Landscape spirit includes feng shui, festivals and rituals, and symbolic images of the landscape, etc. [16].

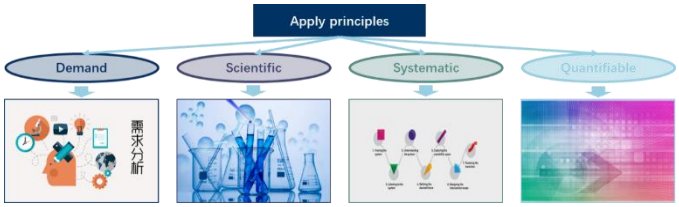


Fig. 4. Principles of application.

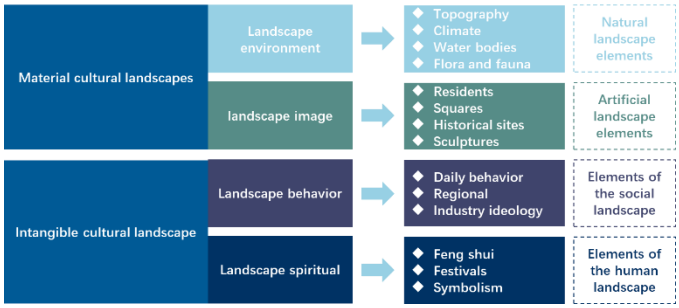


Fig. 5. Extraction of application analysis metrics.

C. Application Analysis and Feasibility Assessment Programme

By analysing the application of VR in the cultural landscape promotion platform, extracting the application analysis indexes, and after a series of data processing, the feasibility assessment model is constructed by adopting the feasibility assessment method of the cultural landscape promotion platform with improved data-driven algorithms to assess the feasibility of the application of virtual reality technology in the cultural landscape promotion platform, and the design of the scheme is specifically shown in Fig. 6. As can be seen from Fig. 6, the feasibility assessment study on the application of virtual reality technology in the cultural landscape promotion platform mainly includes key technologies such as the application analysis of VR technology, the collection of data processing, the construction of the feasibility assessment model, and performance analysis.

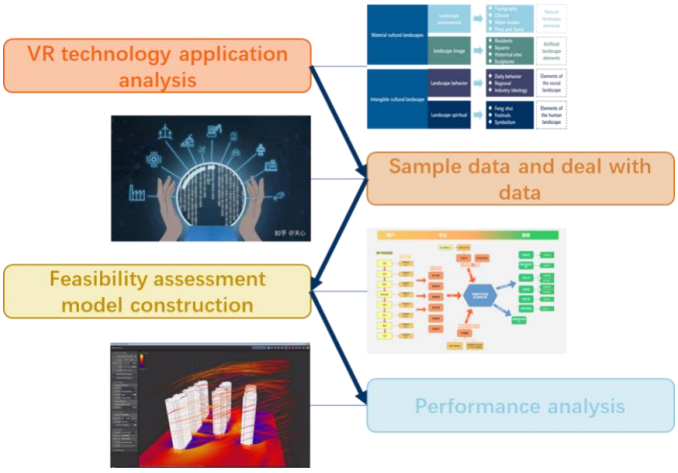


Fig. 6. Feasibility analysis assessment scenario for the application of virtual reality technology in a cultural landscape promotion platform.

III. Feasibility Assessment Algorithm for Virtual Reality Cultural Landscape Promotion Platform based on MA-BiGRU Model

A. Mayfly Algorithm

Mayfly Optimization Algorithm (MA) [17] is an optimization algorithm that mimics the flight and mating behaviour of mayflies in nature. The Mayfly Algorithm is characterised by the fact that its population is divided into two parts: male mayflies move through the global optimum and their own historical optimum positions, while female mayflies move towards a mate that is superior to them, and perform a local search on their own if the mate is weaker than them. The algorithm generates new individuals by simulating the mating process of mayflies and retains the best individuals in the group through a selection mechanism [18].

1) MA algorithm optimisation strategy: The core optimisation strategies of the MA algorithm include the following:

a) Male Mayfly movement: The position update of male mayflies relies on their own speed and the experience of neighbouring individuals, and considering the specific behaviour of mayflies on the water surface, their speed update includes the influence of cognitive and social components, as shown in the following update model:

$$x_i^{k+1} = x_i^k + v_i^{k+1} \tag{1}$$

$$v_{ij}^{k+1} = g \cdot v_{ij}^k + a_1 e^{-\beta r_p^2}\left(p_{best_{ij}} - x_{ij}^k\right) + a_2 e^{-\beta r_g^2}\left(g_{best_{ij}} - x_{ij}^k\right) \tag{2}$$

$$p_{best_{ij}} = \begin{cases} x_i^{k+1} & f\left(x_i^{k+1}\right) < f\left(p_{best_i}\right) \\ p_{best_i} & f\left(x_i^{k+1}\right) \geq f\left(p_{best_i}\right) \end{cases} \tag{3}$$

$$g_{best} = \min\left\{ f\left(p_{best1}\right), f\left(p_{best2}\right), \cdots, f\left(p_{bestN}\right)\right\} \quad (4)$$

where x_i is the position of the ith male mayfly; v_i is the velocity of the ith male mayfly; k is the number of iterations; a_1 and are positive attraction coefficients, which measure the effect of experience gained from self and society, respectively, on the behaviour; g is the coefficient of gravitational force, which decreases with the number of iterations; $p_{best_{ij}}$ is the individual optimum; $g_{best_{ij}}$ is the global optimum; N is the total number of male mayflies in the swarm; r_g is the Cartesian distance between x_i and g_{best} ; r_p is the Cartesian distance between x_i and p_{best_i} :

$$\left\| x_i - X_i \right\| = \sqrt{\sum_{j=1}^{n} \left(x_{ij} - X_{ij} \right)^2} \quad (5)$$

The best male mayflies perform a unique nuptial dance step, moving up and down at different speeds. These male mayflies have a speed of:

$$v_i^{k+1} = v_i^k + d' \times r \quad (6)$$

where d' is the wedding dance coefficient and r is a random number between 0 and 1.

b) Female Mayfly movement: Female mayflies select mates during mating, and their positional updates depend on interactions with male mayflies. The velocity update of female mayflies takes into account attractive and random wandering factors to model their movement through the air. The position of the ith female mayfly is updated as:

$$y_i^{k+1} = y_i^k + v_i^{k+1} \quad (7)$$

Female mayflies and male mayflies are attracted to each other according to the principle of equal ranking of individual fitness, the position of female mayflies changes with the position of male mayflies with the same ranking, i.e.

$$v_i^{k+1} = \begin{cases} gv_{ij}^k + a_2 e^{-\beta r_{mf}^2} \left(x_{ij}^k - y_{ij}^k \right) & f\left(y_i\right) > f\left(x_i\right) \\ gv_{ij}^k + f_{f1} \times r & f\left(y_i\right) \le f\left(x_i\right) \end{cases} \quad (8)$$

where r_{mf} is the Euclidean distance between the ith female mayfly and the ith male mayfly; and f_{f1} is a random wandering coefficient used when the female mayfly is not attracted to the male mayfly.

c) Mayfly mating: The mating process involves the selection of male and female mayflies, usually based on their respective fitness values. Mating produces offspring that replace certain individuals in the population to maintain population diversity and facilitate the search process. The position of mated offspring is calculated as follows:

$$\begin{cases} o_1 = L \cdot m_{male} + \left(1 - L\right) \cdot f_{female} \\ o_2 = L \cdot f_{female} + \left(1 - L\right) \cdot m_{male} \end{cases} \quad (9)$$

where L denotes a random value; o_1 and o_2 initial values are set to 0; m_{male} and f_{female} are the position matrices of males and females, respectively.

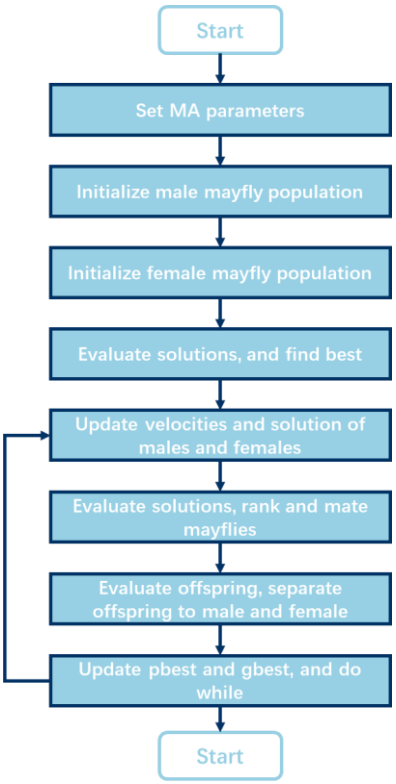


Fig. 7. Flowchart of MA algorithm.

d) Wedding dance and random wanderings: Wedding dance steps and random wandering make the algorithm search locally further, but fixed coefficients may make the mayfly iterate to a worse position later in the iteration. Therefore, make the dance step coefficients and random wandering coefficients decrease with the number of iterations, and update the arithmetic as:

$$\begin{cases} d_t = d_0 \delta \\ f_{fl_t} = f_{fl_0} \delta \end{cases} \quad (10)$$

Where, δ is a random factor with a value between 0 and 1.

e) Offspring variation: To deal with the case of premature convergence, so that the algorithm searches new regions of the space that have never been visited before, the offspring are randomly selected for the perturbed variant of the normal distribution, i.e.

$$o_n' = o_n + \sigma N_n(0,1) \tag{11}$$

where σ is the standard deviation of the normal distribution and $N_n(0,1)$ is the standard normal distribution.

2) MA algorithm flow: According to the optimisation strategy of MA algorithm, the flow of MA algorithm is shown in Fig. 7.

3) MA algorithm application: The mayfly algorithm can be applied to a wide range of optimisation problems, including engineering design [19], machine learning parameter optimisation [20], path planning [21], image segmentation [22] and other fields (Fig. 8). The strength of the algorithm lies in its ability to explore the search space efficiently and to improve search efficiency by simulating biological behaviour [19].

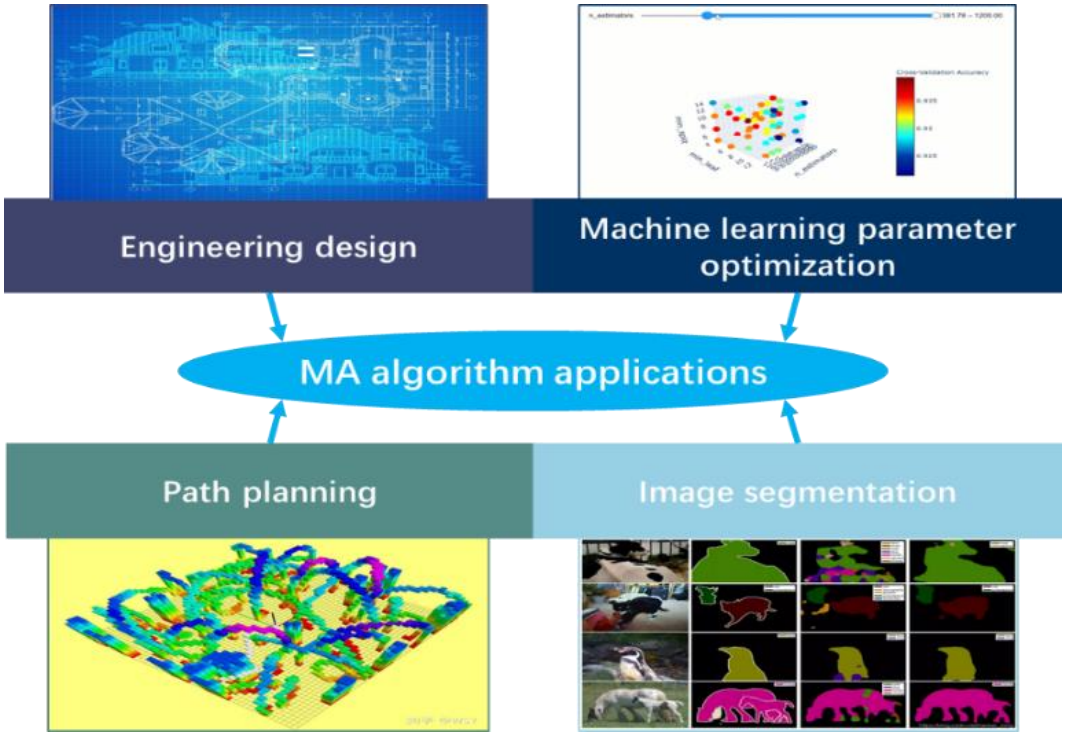


Fig. 8. Application of MA algorithm.

B. The MA-BiGRU Model

1) BiGRU network

a) BiGRU network fundamentals: BiGRU (Bidirectional Gated Recurrent Unit) network [23] is a special form of Recurrent Neural Network (RNN) that combines two directional GRU (Gated Recurrent Unit) layers to process sequence data. Compared to unidirectional RNNs, BiGRU is able to take into account both past and future information of sequence data to better understand contextual dependencies in the data, and the structure is shown in Fig. 9. The output of each time step depends not only on the information before that time point, but also on the information after, which makes BiGRU very useful in tasks such as text analysis, speech recognition, and natural language processing.

b) Characteristics of BiGRU: According to the structural principles of BiGRU network, its characteristics are as follows: i) Bidirectional structure: BiGRU consists of two GRU layers, one processing left-to-right information and the other processing right-to-left information; ii) Gating mechanism: the GRU unit contains reset gates and update gates, which control the flow of information and help the network to capture long-range dependencies; iii) Reducing the gradient vanishing problem: in comparison with the traditional RNN, GRU mitigates the gradient vanishing problem through the gating mechanism, enabling the network to learn long-distance dependencies more efficiently; iv) Parameter efficiency: although BiGRU is able to capture bi-directional information, it usually has fewer parameters than BiLSTM (Bidirectional Long Short-Term Memory), which may result in faster training speeds and lower memory requirements, as shown in Fig. 10.

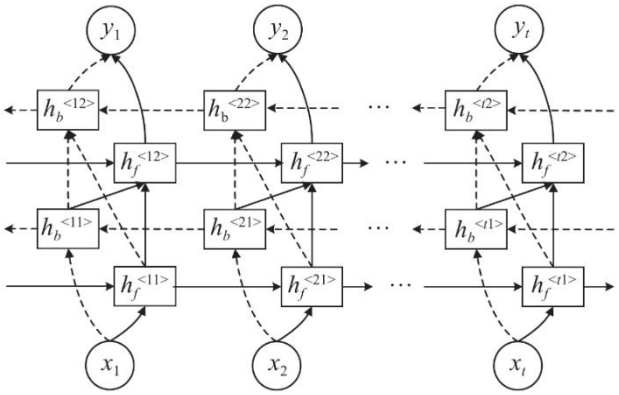


Fig. 9. Structure of the BiGRU model.

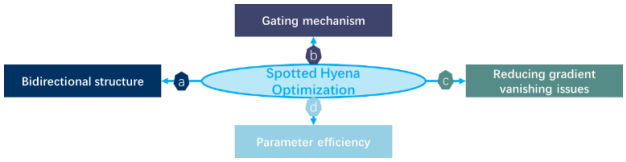


Fig. 10. BiGRU characteristics.

c) Application of BiGRU: BiGRU network is widely used in a variety of deep learning tasks due to its ability to process sequence data efficiently. For example, it can be used in natural language processing tasks such as text sentiment analysis, named entity recognition, and machine translation, as well as in the fields of time series prediction and speech recognition [24] (Fig. 11).

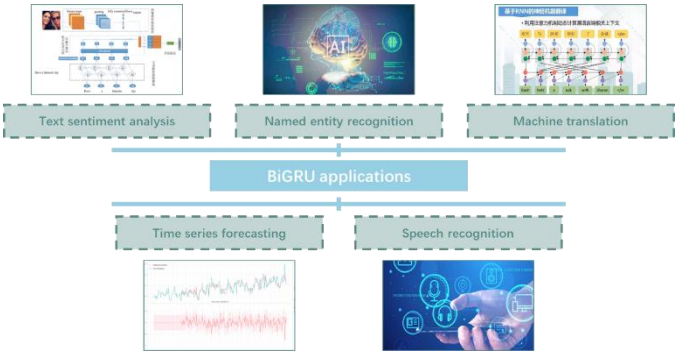


Fig. 11. BiGRU application.

2) MA-BiGRU model

a) Optimising variable coding: In the optimisation variables section, this paper uses real number coding for Update Gate, Reset Gate and Candidate State Weights and Bias of BiGRU network as shown in Fig. 12.

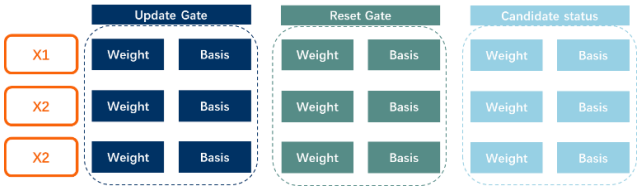


Fig. 12. Optimising the coding structure of decision variables.

b) Adaptation function setting: The MA algorithm is introduced mainly to improve the mapping accuracy of BiGRU model, so in this paper, RMSE, MAPE, MAE, R2 are selected as the fitness function in the optimisation process of MA algorithm (Fig. 13), and the specific fitness function is calculated as follows:

$$F(x) = w_1 f_{RMSE} + w_2 f_{MAPE} + w_3 f_{MAE} + w_4 (1/f_{RMSE}) \quad (12)$$

where w_1 , w_2 , w_3 and w_4 are the weights of RMSE, MAPE, MAE, and R2 , respectively.

Fig. 13. Structure of the fitness function.

c) Analysis of model steps: According to the decision variable coding and fitness function settings, the MA-BiGRU model is constructed by determining the parameters of the MA algorithm, generating the initial population, updating the update gates, reset gates, and candidate state weights and biases of the BiGRU network by using the optimisation strategy of the MA algorithm, comparing the values of the fitness function, and obtaining the optimal update gates, reset gates, and candidate state weights and biases of the BiGRU network, and thus constructing the MA-BiGRU model, and the pseudo-code is shown in Fig. 14. Fig. 15 shows the schematic diagram of SHO-LSTM structure.

Algorithm 1: BiGRU based on MA	
1	Determine optimized variables, including weights and basis;
2	Set MA algorithm parameters;
3	Encode mayfly population;
4	Calculate fitness, and update best mayfly;
5	Do While stop criteria are not met
6	Update velocities and solutions of males and females;
7	Evaluate solutions;
8	Rank the mayflies;
9	Mate the mayflies;
10	Evaluate offspring;
11	Separate offspring to male and female randomly;
12	Replace worst solutions with the best new ones;
13	Update *pbest* and *gbest*;
14	End While
15	Output best parameters of BiGRU;
16	Build MA-BiGRU model.

Fig. 14. MA-BiGRU pseudo-code.

Fig. 15. Schematic diagram of SHO-LSTM structure.

IV. Example Analyses

A. (1) Introduction of Examples

In order to verify the application of VR technology in cultural landscape promotion platform analysis and usability assessment algorithm effectiveness, this paper takes Dongao Ding Village in Dongtou District, Wenzhou City as a case study. Dongao Ding Village is located in the southeast end of Dongtou Island. Dongao Ding village coastline reefs are strange and strange, and the sandy beach is broad and soft, which has an important economic location advantage (Fig. 16).

Site location

To the east lies Senile rock、And half screen across the sea,Jingyuan has obvious advantages Dongao Ding Village is a coastal fishing village in the southeast of Dongtou Island, facing Banping Mountain across the sea. Dashatun and Xiandiengyan scenic spots are located in the village, with obvious advantages of scenic source, and are located in the important position of "Sea Garden - Hundred miles coastal style belt".

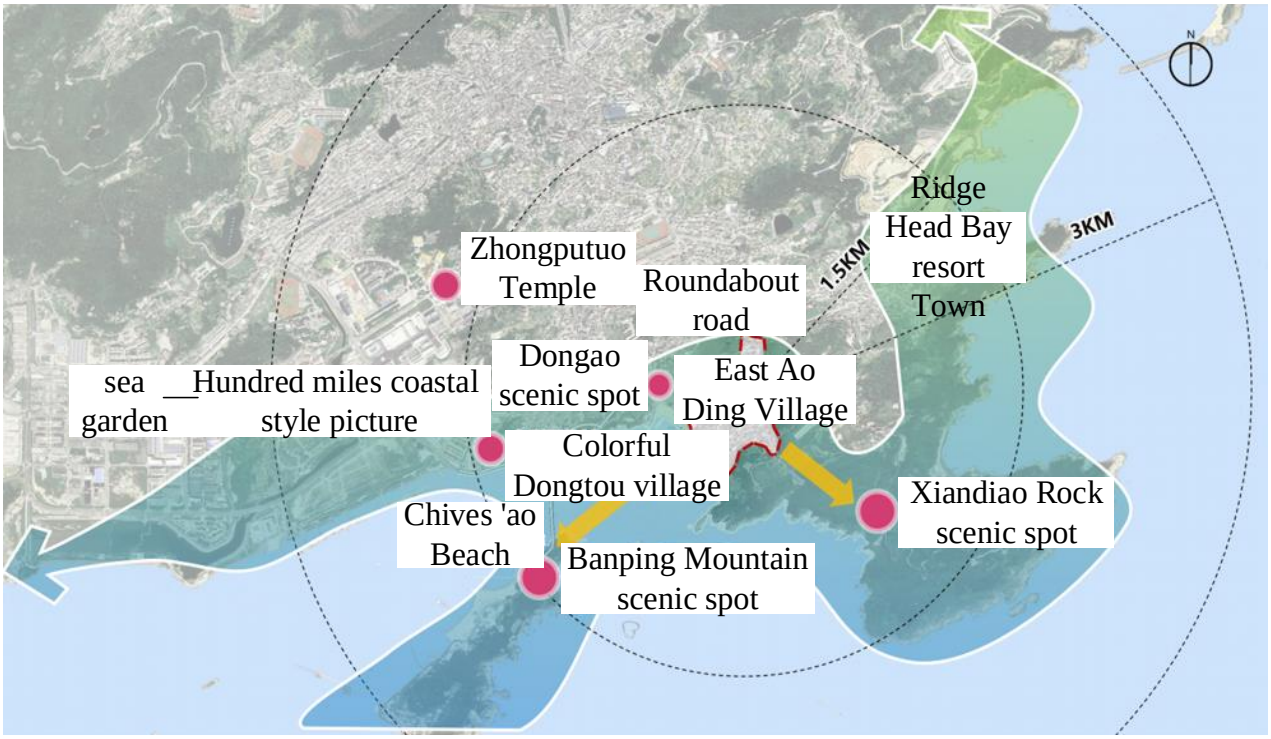


Fig. 16. Dongao Ding Village, Dongtou District, Wenzhou City, China.

In the usability evaluation algorithm experiments, this paper adopts LSTM, GRU, BiGRU, MA-BiGRU models as the analysis and comparison algorithms; LSTM, GRU, BiGRU are optimised iteratively using Adam's algorithm, the activation function is the Relu function, and the number of nodes in the hidden layer is 50; the NP of MA algorithm is set to 100, and the Max_iteration is set to 1000.

B. Example Analysis

1) Application results: Taking the landscape of Dongao Teng Village as an example, this paper uses VR technology in the cultural landscape promotion platform application analysis to extract the landscape elements of various cultural landscapes, and the specific results are shown in Fig. 17.

Fig. 17. Cultural Landscape Elements of Dongao Ding Village.

Fig. 18 gives the effect of promoting the historical and cultural landscape of Dongao Ding Village. Fig. 19 gives the effect of virtual reproduction of cultural landscape based on VR technology.

Fig. 18. Historical and Cultural Landscape Promotion Effect of Dongao Ding Village.

Fig. 19. Effect of virtual reproduction of the historical and cultural landscape of Dongao Ding Village.

2) Usability evaluation algorithm analysis: In order to test the effectiveness of the usability evaluation algorithm, this paper uses 73 samples for training and tests on samples 73-80, and the analysis results are shown in Fig. 20 and Fig. 21. Fig. 20 demonstrates the comparison of the evaluation level results of the usability evaluation algorithms based on LSTM, GRU, BiGRU, and MA-BiGRU models.

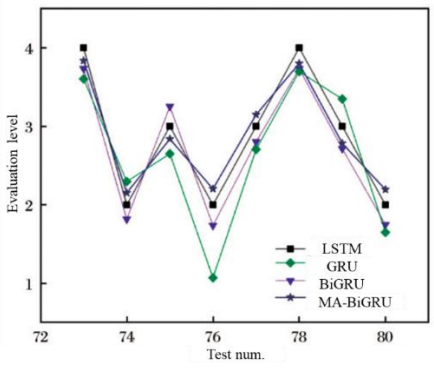


Fig. 20. Assessment level results.

The absolute errors of the usability evaluation algorithms based on LSTM, GRU, BiGRU, and MA-BiGRU models are plotted in Fig. 21. From Fig. 21, it can be seen that MA-BiGRU has the smallest absolute error.

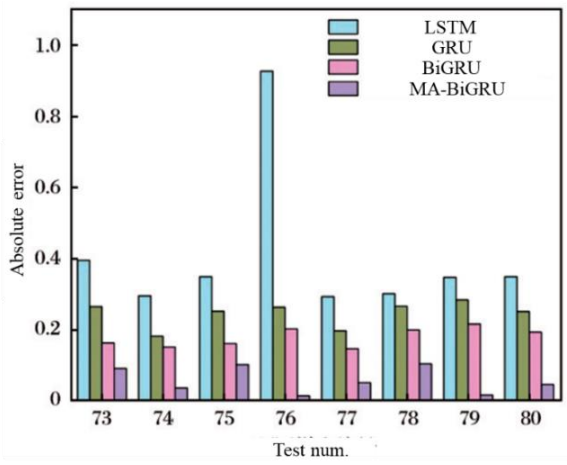


Fig. 21. Comparison of absolute error results.

V. Conclusion and Outlook

This paper proposes an application analysis and usability assessment algorithm of VR technology in cultural landscape promotion platform based on MA algorithm and BiGRU network. Taking the application of VR technology in the cultural landscape promotion platform as the research object, analysing the application situation, designing the application feasibility assessment scheme, and combining the MA algorithm and BiGRU network, an application analysis and usability assessment method based on the MA-BiGRU model is proposed. Taking Dongao Ding Village as a case study for analysis, the results show that the assessment method based on MA-BiGRU model has the best assessment and prediction accuracy and can improve the application effect.

Despite these promising results, the study has some limitations. First, the focus on a single case study limits the generalizability of the findings. Second, the computational complexity of the MA-BiGRU model may pose challenges for real-time usability in larger-scale projects.

Future research could explore three key directions: 1) Expanding the usability model to encompass a broader range of cultural landscapes across different regions. 2) Developing a more computationally efficient version of the MA-BiGRU model to facilitate real-time usability. 3) Integrating other machine learning models or algorithms to enhance the robustness and flexibility of the assessment framework. By addressing these areas, future work can further refine VR technology's role in cultural landscape promotion and its practical applications.

Acknowledgment

This work is supported by Zhejiang Provincial Science and Technology Innovation Program (New Young Talent Program for College Students). Research and Design of a One-Cloud Multi-End Rural Service Platform Based on the Background of Cultural and Tourism Integration. Grant No.: 2024R429B055.

References

[1] A A , Xu M , Gonalez-Redin J , Ali A, Shahzad L, Rahim S.Spatiotemporal valuation of cultural and natural landscapes contributing to Pakistan's cultural ecosystem services[J].Environmental Science and Pollution Research, 2022, 29(27):41834-41848.

[2] Fontefrancesco M , Zocchi D M , Pieroni A .The Intersections between Food and Cultural Landscape: Insights from Three Mountain Case Studies[J]. 2023.

[3] Carroll J .Peter Handke's Landscapes of Discourse. an Exploration of Narrative and Cultural Space by Christoph Parry (review)[J].Austrian Studies , 2022, 13:284 - 285.

[4] Gao C , Iqbal J .An empirical study of Thai cities' colour landscapes[J].Heliyon, 2023, 9.

[5] Cai Xiaoyao,Jin Yuanli.Example of the use of VR virtual reality technology in the study of cultural landscape protection and inheritance of ancient towns[J]. China Building Metal Structure,2022,(06):105-107.

[6] Li Shiyan, Wang Lu. Application and thinking of virtual reality technology in cultural theme park--Taking the conceptual plan of "Jin Yong martial arts theme park landscape design" as an example[J]. Modern Horticulture,2022,45(07):133-135.

[7] Mo Wenshui. Reproduction of folk culture landscape based on virtual reality technology[J]. Culture Industry,2022,(27):13-15.

[8] Yang Jie, Sun Danfeng. Research on the new situation of cultural communication and tourism business promotion of historical and humanistic landscapes--Taking the virtual reality display of Nanjing Fuzimiao as an example[J]. Art Technology,2016,29(02):134.

[9] Dimitar S , Alexander G , Dimka V S .Kamchatka-the Cold and the Heat of the Earth[J].Geoheritage, 2023, 15(4).

[10] Xu Y , Xie M , Xiong S .Application of Virtual Reality Technology in Historical and Cultural Landscape Reproduction[J].Applied Mathematics and Nonlinear Sciences, 2024, 9(1).

[11] H.L. Wu,S.Liu. Mayfly algorithm for elite coevolution[J]. Journal of Zhejiang University (Engineering Edition),2024,58(07):1346-1356.

[12] He J , Zhang S , Fang C .AAindex-PPII: Predicting polyproline type II helix structure based on amino acid indexes with an improved BiGRU-TextCNN model [J].Journal of Bioinformatics and Computational Biology, 2023, 21(05).

[13] Bezáková, Magdaléna,Bezák, Peter.Which sustainability objectives are difficult to achieve? The mid-term evaluation of predicted scenarios in remote mountain agricultural landscapes in Slovakia[J].Land Use Policy, 2022, 115.

[14] Kikuchi Y .Residents' Perceptions of Viticultural Landscapes[J].Reports of the City Planning Institute of Japan, 2023, 21(4):432-438.

[15] Xiangyun X , Xiaoshuang Y , Shizhen X , Dijin M U. Spatial Distribution and Inscribed Criteria of Cultural Landscape World Heritage Sites[J]. Landscape Research:English Edition, 2023, 15(1):63-67.

[16] Mora F G S D L .The Cultural Landscapes of Maya Roads: The Material Evidence and a GIS Study from the Maya Lowlands of Chiapas and Tabasco, Mexico[J]. Latin American Antiquity, 2023.

[17] Zervoudakis K, Tsafarakis S. A mayfly optimisation algorithm[J]. Computers & Industrial Engineering, 2020, 145: 106559.

[18] CHEN Tao,LI Xin. Application of terahertz spectroscopy in the identification of genetically modified rapeseed oil:a support vector machine model based on an improved mayfly algorithm[J]. Journal of Physics,2024,73(05):366-374.

[19] XI Tao, GE Zengyuan, WANG Lijing. Fault diagnosis of rolling bearing based on parameter optimised VMD and DBN[J]. Combined machine tools and automated machining technology,2022,(12):57-61.

[20] YIN Huiyan, CHANG Yong, YANG Hailan, WU Wencheng, KUNG Kai. Short-term load forecasting method based on MA-LSTM[J]. Electrotechnology,2023,(21):15-19.DOI:10.19768/j.cnki.dgjs.2023.21.004.

[21] XI Wanqiang, CHANG Baoshuai, LIN Siwei, LIN Junzhi, LI Peng. Multi-strategy improved mayfly algorithm for UAV trajectory planning[J]. Electro-Optics and Control,2023,30(11):80-84.

[22] He Hang, Xu Lianjie, Li Gaoyuan, Lv Rongfei, Wang Xiliang. Optimising multi-threshold image segmentation based on improved mayfly algorithm[J]. Science Technology and Engineering,2024,24(12):5059-5068.

[23] Yang H , Zhang Z , Zhang L .Network security situation assessments with parallel feature extraction and an improved BiGRU[J]. University (Science and Technology), 2022, 62(5):842-848.

[24] Liu W , Gu Y , Ge Y .Multi-factor stock trading strategy based on DQN with multi-BiGRU and multi-head ProbSparse self-attention[J].Applied Intelligence, 2024, 54(7):5417-5440.

Migrating from Monolithic to Microservice Architectures: A Systematic Literature Review

Hossam Hassan, Manal A. Abdel-Fattah, Wael Mohamed
Information Systems Department, Helwan University, Egypt

***Abstract*—Migration from monolithic software systems to modern microservice architecture is a critical process for enhancing software systems' scalability, maintainability, and performance. This study conducted a systematic literature review to explore the various methodologies, techniques, and algorithms used in the migration of monolithic systems to modern microservice architectures. Furthermore, this study underscored the role of artificial intelligence in enhancing the efficiency and effectiveness of the migration process by examining recent literature to identify significant patterns, challenges, and optimal solutions. In addition, it emphasizes the importance of migrating monolithic systems into microservices by synthesizing various research studies that enable greater flexibility, fault tolerance, and independent scalability. The findings offer valuable insights for both researchers and practitioners in the software industry. In addition, it provides practical guidance on implementing AI-driven methodologies in software architecture evolution. Finally, we highlight future research directions in providing an automation technique for the software architecture migration.**

***Keywords*—Software migration; software evolution; monolithic architecture; microservice architecture; systematic literature review**

I. Introduction

Over the last decade, software architecture has significantly changed. Software was initially monolithic, integrating all components into a single unit. However, as software systems grew more complex, limitations in scalability, flexibility, and maintainability appeared. This encourages an investigation of intermediary architectural patterns like micro-kernel architecture and service-oriented architecture (SOA). Micro-kernel architecture has a minimal core system with only the most important functions while delegating additional functionalities to modular, user-space components, whereas SOA focuses on breaking software down into loosely coupled services. These evolutions paved the way for a new microservices architecture, which divides software into tiny, autonomous, single-need components.

Most software businesses worldwide have adopted the modern microservice architecture because of its significantly increased scalability, maintainability, flexibility, and ease of development [1], [2]. In addition, many software businesses, like Amazon, Uber, Netflix, and Spotify, are adopting this architectural approach, and the transition to microservices is well underway [3]. Microservice architecture can be represented by a collection of tiny services, each operating in its own process and interacting using lightweight protocols like HTTP (Hypertext Transfer Protocol), developed around business needs and delivered independently [2], [4].

The microservices architecture solves many monolithic system issues. Some of the major benefits of microservices include scalability, which improves resource utilization and peak performance, and flexibility, which allows development teams to pick the best tools and technologies for each service, resulting in more customized and efficient solutions, as well as fault tolerance, which prevents system failures from affecting all services, improving resilience and reliability [3].

During the migration to the microservice architecture, some software warehouses that have monolith-based systems attempt to decompose them into coherent microservice-based implementations. The purpose of this decomposition is to occasionally assist software architects in identifying microservice candidates by analyzing the application's domain, business needs, source code, and version-related information [5], [6], [7].

Nevertheless, software vendors have been exerting efforts to manually migrate from monolithic to microservice-oriented application ecosystems. Scalability, component independence, data management, service communications, deployment, and monitoring are some of the efforts [1], [8]. This migration process is subjective, requires human judgment, and is prone to errors. Expert opinions are required for this process, as is software engineers' proficiency in microservice extraction and system quality preservation [9]. Besides, extracting microservices is becoming a complex process because there is no clear or straightforward method for defining the boundaries of microservices. The first and most challenging stage in breaking down a monolithic application involves identifying microservice boundaries. Insufficient identification may lead to more complex systems with lower quality [1].

To address this gap, there is an increasing interest in using AI and ML algorithms to facilitate the migration process. AI-driven methods, such as search-based techniques and clustering algorithms, make it possible to automatically find microservice components and improve the decomposition of systems. However, the majority of the studies fail to identify relevant and potential microservices, and they struggle to determine the appropriate number of candidate microservices while also ensuring their granularity and loose coupling [10].

This paper aims to conduct a comprehensive and systematic literature review to analyze and identify the most important methods and techniques used for facilitating the migration process of software from monolithic systems to modern microservices. The review also aims to compare the results in a detailed manner through a systematic literature review (SLR),

which can serve as a foundation for developing effective solutions.

This paper's subsequent sections follow this structure: Section II encompasses the background and motivation, while Section III presents the research methodology. Section IV provides an overview of the current state of knowledge and understanding regarding the process of software migration. Section V presents the findings of this literature review. Section VI covers the conclusion of this literature review and offers suggestions for future work and improvements.

II. Background

Monolithic and microservice architectures are two different methodologies for designing and constructing software systems. A variety of factors influence the choice between them. Each architecture possesses its own advantages and disadvantages, and the determination should be made considering the particular requirements and limitations of the project. In the following subsections, we will provide additional information regarding the similarities, differences, advantages, and disadvantages of the two architectural styles.

A. Monolithic Architecture

A single deployed unit is referred to as a monolith. A monolithic system necessitates the simultaneous deployment of all functionalities. Deploying all code as a unified process, consolidating all code into a single process, characterizes a monolithic architecture [2].

The modular monolith, a new version of the single-process monolith, divides the single process into multiple distinct modules developed separately; however, deployment requires the combination of these modules. It might be an optimal solution for several enterprises and software warehouses because it defines module boundaries well, provides a significant amount of parallel work, overcomes the complexities associated with the distributed microservice architecture, and adopts a simpler deployment topology [2], [11]. Differences between monolithic and modular monolithic software architectures can be shown in Fig. 1. Shopify is an example of a company that uses this technique as a substitute for microservice deconstruction, and it appears to be quite effective for them.

Unfortunately, the monolith has recently become a symbol of avoidance and is often associated with legacy systems. However, it is actually a viable option based on the system's requirements and specifications. Some of the advantages of monolithic architecture are listed below [2], [12]:

1) Faster and rapid deployment topology: Avoids numerous distributed system issues.

2) Workflows for developers are simpler to manage.

3) Simpler monitoring and troubleshooting: End-to-end testing is simplified.

4) Code reuse is simple enough, without any duplication.

Nevertheless, an important issue with the modular monolith architecture is that the database does not have the same degree of decomposition as the code, making it difficult to separate the monolith in the future [2].

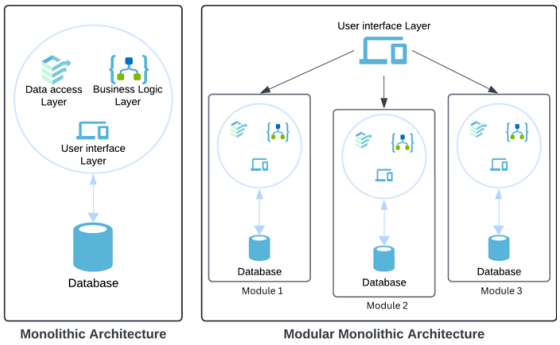


Fig. 1. Demonstrating the distinctions between monolithic and modular software architecture.

B. Microservice Architecture

Microservices are designed to be loosely coupled and independently deployable, but they may still rely on different coordination patterns to handle distributed transactions and inter-service communication, which are designed based on a specific business domain and potentially released separately. A service is a module that has specific functionality and allows other services to access it over networks. By combining these modules, the software warehouses could create more intricate and enterprise systems. Each microservice might represent a specific aspect of the system. When combined, these microservices form a complete enterprise system. In other words, they are a kind of service-oriented architecture that has a certain viewpoint on how service boundaries should be defined with the ability to deploy independently. Fig. 2 showed an example of a microservice architecture construction.

A single microservice appears as a black box. One or more network endpoints, such as a SOAP or REST API, host business functionality using appropriate protocols. Through these networked endpoints, consumers—microservices or programs—access this capability. The outer world conceals implementation elements like service technology and data storage. In most cases, microservice designs encapsulate their own databases instead of using common databases.

Microservices conceal information within their components and deliver minimal information through external interfaces. The hidden microservice implementation can be freely updated as long as the changes do not introduce incompatible modifications to the network interfaces it exposes. Changes made within a microservice boundary should not affect upstream consumers, allowing for separate functionality releases. Clear, consistent service boundaries that don't alter with internal implementation lead to looser coupling and greater cohesiveness. Some of the advantages of microservice architecture are listed below:

1) Deployability independence: Allows for modification and deployment without relying on other microservices.

2) Business domain-based model: Makes it easy to bring out new functionality and recombine microservices to provide consumers with new capabilities.

3) Owning their own state: Microservices must avoid relying on shared databases. Instead, it needs to request data from another microservice in order to access it.

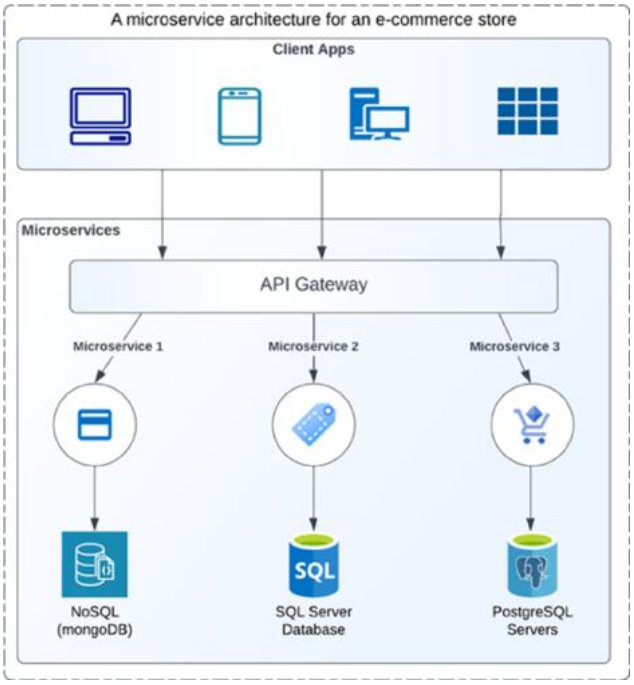


Fig. 2. A microservice architecture for an online shopping system.

III. Research Methodology

To achieve our goal of automatically migrating software architecture from monolithic to microservice, SLR was employed to analyze the migration process. This investigation encompassed the use of AI and other techniques to automate the process. SLR was chosen in this study due to its advantages over conventional literature reviews in terms of accuracy, effectiveness, and organization. It aids in the identification of our objectives, the assessment of their outcomes, and the classification of them into categories. This SLR is comprised of three primary steps, which are as follows [13], [14], and [15]:

1) Planning and preparing the review.
2) Performing the review.
3) Reporting the review.

A. Phase 1: Review Planning

Starting with these steps, this phase covers the core of SLR:

1) Clearly state questions the study will address:

RQ1: Which AI methods are effective for automated microservice architecture migration?

RQ2: What are the primary obstacles and impediments that organizations face during the migration process, and how can AI-driven solutions overcome these challenges?

RQ3: What are the criteria used to assess the effectiveness and success of AI-driven migration strategies?

2) Choose appropriate research repositories: Within the SLR, various online digital libraries were used, including but not limited to IEEE Xplore, Google Scholar, Science Direct, Springer, and MDPI.

3) Establishing research criteria: The search strings were chosen based on the SLR keywords and the alternatives to those keywords found in Software Architectural Evolution. As can be seen in Table I, these search strings were divided into two distinct categories.

4) Software Migration Process.
5) Monolithic to Microservice.

TABLE I. Search Strings for the Systematic Literature Review

GROUP	Search String
Software Migration	(“Software Migration” OR “Software Architectural Evolution” OR “Software Architecture Transformation” OR “Software Evolution” OR "Software Architecture")
	AND
monolithic to Microservice	(("Legacy", "Monolithic", "Monolithic system", "Single-layer application", "Modular Monolithic")) AND (("Microservices", "Microservices pattern", "Microservice architecture", "Service-oriented architecture (SOA)")

6) Providing clear definitions of inclusion & exclusion: The guidelines for the inclusion and exclusion criteria for this SLR were established by Kitchenhem [13].

The following criteria for inclusion were listed:

IC1: A journal publication or conference presentation are required for the research selection.

IC2: Studies should focus on software migration process, especially the migration from monolithic to microservice.

IC3: AI, ML, and other algorithms or solutions for software migration should be applied.

IC4: Studies published between 2018 and 2024. Microservices were introduced earlier, but the latest studies show the recent implementations, challenges, and innovations.

The following criteria for exclusion were listed:

EC1: Studies that failed to address the research questions.

EC2: Studies that ignored software migration process.

EC3: Studies that didn’t include microservices.

EC4: Publications published prior to 2018.

7) Establishing standards for measuring quality: This phase was significant because we compared and checked the quality of the selected studies to our objectives, research questions, and goals.

B. Phase 2: Performing the Review:

1) Primary data selection: During this stage, filtration methods were employed to apply search criteria and determine inclusion and exclusion criteria. The process of selecting the primary studies has started. The Tollgate approach was implemented to enhance the effectiveness and efficiency of the selection process in a systematic and organized manner [13].

2) Data extraction: The study selection was guided by specific criteria, including research methodology, publication

year, kind of study, and any restrictions or limitations imposed on the studies.

3) Data synthesis: The collected studies were assessed and compared with our research questions and study objectives.

C. Phase 3: Reporting the Review:

During this phase, selected studies were verified and compared against quality criteria. Fig. 3 shows the process of SLR, indicating a well-organized collection of studies available for discussion and investigation, with R representing the number of research studies in each step.

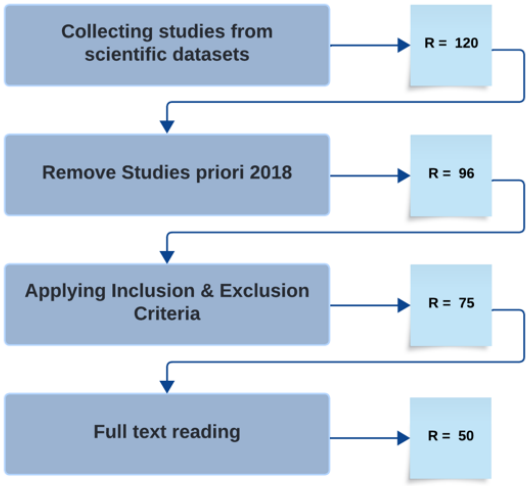


Fig. 3. The SLR selection process.

IV. State-of-the-Art

This section examined the studies selected for analysis, reviewing, and discussion.

Jin et al. [16] proposed a functionally-oriented microservice extraction (FoME) method that uses clustering of execution traces and classifies source code entities according to their functionality. The authors evaluated their method against three methods (LIMBO, WCA, and MEM) across four open-source projects. Their findings showed that FoME produces microservices with comparable cohesiveness to other methods and achieves much looser coupling. However, their method relies on high-quality test cases, with future efforts geared toward full automation of the process.

Sellami et al. [1] proposed a method called MSExtractor, which identifies microservices as multi-objective optimization problems using an indicator-based evolutionary algorithm (IBEA) while considering structural and semantic dependencies in the source code. They conducted a benchmark of seven software systems to assess the effectiveness of their method. Their findings demonstrated that MSExtractor outperformed other clustering algorithms like FoME and MEM mentioned in [16], [17]. Nevertheless, their method was limited to four web applications and lacked generalization. Future work should consider non-functional evaluation metrics.

Velepucha et al. [18] concentrated on breaking down microservice architectures. They compared different concepts, compiled a list of microservice architecture patterns as shown in Table II, and discussed the advantages and disadvantages of microservices over monolithic architectures. Future work and limitations include adapting micro-frontends, automatically migrating the decomposition process, restricting the results to an object-oriented approach, and further evaluating the literature.

TABLE II. List of Patterns used in Microservice Decomposition Process

Pattern / Description
Domain-Driven: Developing software systems focusing on business logic.
Service discovery: Addressing service interactions and communications.
Data-driven: Design systems around data behavior or structure. Data is key.
Backend for frontend: Developing services tailored to frontend clients.
Adapter microservice: Transforming microservices functionality data.
Strangler-application: Incrementally migrate to microservice.
Shared data microservice: Sharing and managing microservice data.
Aggregator microservice: Aggregate data from multiple microservices.

On the other hand, Kazanavičius et al. produced the conceptual model, which aims to migrate a monolith database into a multi-model polyglot persistence system [19]. They assessed their proof of concepts using the ISO/IEC 25012:2008 standard's definition of quality attributes. Their findings indicate that their proposed method can effectively transition data storage from a monolithic to a microservice architecture. Future direction includes the need to automate the solution and the fact that their model's adoption depends on single app.

Nordli et al. [20] focused on monolithic solution vendors; they struggle to convert monolithic products into multi-tenant cloud-native SaaS solutions because many clients, especially large enterprises, want customized products. The authors presented a proof of concept that outlines a combined approach for transforming monoliths into microservices-based, customizable cloud-native SaaS. Their findings demonstrated that a customization-driven migration approach can guide a monolith towards becoming a SaaS. Their limitations included the unavailability of datasets, the need for expert evaluation and the use of real-world systems to generalize the results, and the requirement for an automatic migration process.

Velepucha et al. [21] performed a SLR and provided a list of challenges, and benefits that arise when carrying out the migration process, as shown in Table III. Some limitations and future work include the necessity to analyze the pros and cons of each architecture, as well as automate the migration process.

TABLE III. Challenges Encountered Throughout the Software Migration

Problem/Challenges	
Utilizing an appropriate tool during the migration process.	Perform the entire microservices migration without breaking it down.
Reorganizing stakeholders is required when implementing microservices.	Challenges in identifying and designing microservices.
Desiring to transition all monolithic programs to microservices	Ensure the consistency when transitioning across databases.
Attempting to integrate new technologies into a monolithic application.	

Faustino et al. [11] performed a case study of transforming systems to microservices using a modular monolithic architecture. They discovered that a modular monolithic architecture could simplify the migration process. Additionally, it addresses issues related to performance optimization, eventual consistency, and inter-microservice communication. However, their method's limitation to a single migration case study hinders its generalizability.

Blinowski et al. [12] compared the performance and scalability of monolithic and microservice architecture by implementing a reference web application with two different technologies and architectural styles. Moreover, they selected three distinct deployment scenarios. Their findings indicated that monolithic architectures outperform microservices on a single machine, with Java handling computation-intensive tasks more efficiently than .NET. For non-computational services on machines with limited power, Azure's vertical scaling proves more cost-effective than horizontal scaling. In future studies, the authors intend to enhance the complexity of their benchmarking system, broaden its application across cloud platforms, and incorporate more performance metrics.

Bastidas Fuertes et al. [22] used Transpiler in the software architecture design model's back-end layer to automatically transform business logic from one source code to several equivalent versions. They tested the model for performance, scalability, and reliability in various scenarios and compared it with existing software design models to identify its pros and cons. Their result showed that the proposed model seeks to save costs, optimize the development process, and enhance the effectiveness of multi-programming language platforms. Nevertheless, the authors acknowledged the need for refining the model's effectiveness for generalizing the result.

Mazzara et al. [8] presented a case study to illustrate the advantages of transforming a monolithic into a microservice on scalability. The case study focuses on the FX Core system, which is critical for Danske Bank. They compared the two architectures, and the results showed that microservices have resulted in enhanced scalability and effectively resolved the significant issues posed by the monolithic. However, their findings lack insights, validation, and verification. Furthermore, the case study utilizes agile methodologies, which limits the generalizability of the results.

Assunção et al. [10] adopted an approach to redesigning features into microservices by employing search-based techniques to quantitatively assess potential redesign possibilities for monolithic features as microservices based on four criteria: coupling, cohesion, network overhead, and feature modularization. To evaluate their findings, an interview with eight monolithic system developers was conducted to get their feedback. Their findings showed that their approach demonstrates positive results and encourages more exploration. To justify configurability, future directions include generalizing results and suggesting specific criteria associated with variability.

Teguh Prasandy et al. [23] presented a method of modularizing application source code, databases, and cloud servers to identify the necessary preparations needed to make a successful transition to microservices. Their findings demonstrated that migrating to microservices can present challenges and affect stockholders, particularly system analysts and developers. Moreover, it's crucial to isolate program blocks during deployment and determine the upload time to prevent any failures. They utilized the Postman tool to assess the REST API as both a REST client and an application. Future research will include assessing the capacity of cloud servers, and evaluation is necessary to generalize the findings.

By calculating the metrics (latency, throughput, scalability, CPU usage, memory usage, and network usage) needed to compare the source and target applications, Fondazione et al. [24] developed a method to determine if migrating to microservices is beneficial. Their findings showed that monolithic works well with small to medium systems, which are typically defined by the project's overall size and complexity. The substantially higher scalability ratio of the microservice system supports the hypothesis that it performs better than a monolithic design for systems with too many concurrent users, especially when it comes to handling more traffic. Additionally, researchers conducted benchmarking experiments to evaluate their results. Future work includes testing on a real-life system, utilizing an alternative programming language, and ensuring security.

Abgaz et al. [25] conducted a SLR by examining 35 studies. Their results showed that the process of breaking down a monolith into microservices is still in its early stages, and there is a lack of techniques for integrating static, dynamic, and evolutionary data. The lack of adequate tool support is also apparent. The author conducted their SLR using Barbara Kitchenham's principles as a guide, as we illustrated in Section III. The authors suggested focusing on microservice deployment and standardizing analysis measurements.

Tapia et al. [26] assessed the performance and correlation of monolithic and microservices applications. They stress-tested their results using the same characteristics and hardware specifications. Furthermore, a mathematical model using the non-parametric regression method verified their studies' findings. Their results showed that monolithic and microservice can serve various technological situations. However, microservices improve hardware resource efficiency, cost savings, and productivity. Future directions include enhancing information security and combating cyberattacks. Moreover, automation tools for migration are required.

Kuryazov et al. [27] proposed a conceptual model for solving the issue of migrating from a monolithic architecture to microservices, especially the decomposition steps. Their conceptual model is still in its early stages and needs more evaluation and testing in a real industry. They need to develop a method for evaluating software using cohesion and coupling measurements, which will simplify the analysis of monolithic systems and estimate the migration effort. Additionally, they should create a tool that facilitates the extraction of software metadata and business logic.

Auer et al. [28] presented a decision support framework for software companies seeking to transition to microservices. Their framework is based on an examination of a set of characteristics and metrics, which they collected and reviewed through interviews with experts. Their findings provided data

and measurements that companies could use to assess microservices adoption. Future work includes validating the framework, identifying automatically applicable measures that can easily reduce decision subjectivity, and adding cloud-native technologies and micro-frontend architecture.

Daoud et al. [7] used a business process to identify needs, data, and semantics to capture dependencies between these processes, as well as a collaborative clustering technique to recommend microservices. Results showed that the approach outperformed similar ones for microservices identification and highlighted the importance of business processes. Future directions are as follows: generating new activity relationships utilizing powerful machine learning, including NLP, and evaluating various activity dependence models for microservice identification. The identification of microservices may be impacted by security concerns, which could be an intriguing development.

Hasan et al. [29] presented a collection of software architecture metrics, including coupling, complexity, cohesion, and size, to assess the maintainability of microservice architectural designs. Results showed that the suggested metrics criteria are more applicable for implementation in industrial settings. On the other hand, case studies from the real-world industrial sector need to be analyzed and applied to the suggested metrics to assess their efficacy. Furthermore, a tool-based methodology must be developed for evaluating the architectural quality of potential microservices.

Oumoussa et al. [30] performed a systematic literature review that highlighted critical areas requiring more attention, such as enhanced automated identification tools and standardized evaluation standards. Their findings showed that many techniques exist for identifying microservices; however, they often focus on particular challenges and abandon others. In addition, there is a dearth of studies that concentrate on a solution to address the migration problem.

Abdellatif et al. [31] conducted a comprehensive analysis of 41 studies. Their research aimed to identify the various inputs, processes, outputs, and usability of service identification methodologies in order to modernize monolithic software. Their findings demonstrate that the categorization aligns with the experiences of industry professionals and provides valuable assistance to practitioners in real-world industrial settings. Future directions include proposing an approach for identifying services based on their types, which enhances the potential for reusing them across several levels: application, enterprise, and business, besides generalizing their results.

Li et al. [32] proposed a technique to identify microservices by utilizing the unified model language (UML), which is derived from the source code. Then, the classes and sequence diagrams were analyzed, using them as input for clustering techniques to identify potential microservices. In addition, experiments were conducted to evaluate the proposed model and compare it to recent methods. Their findings revealed that the proposed model outperformed existing models. However, their results are not generalizable because their proposed model disregards microservice distributions and quality criteria.

The primary focus of Gomes Barbosa et al. [33] was to identify potential microservices from database procedures, specifically targeting monolithic applications developed in the 1980s and 1990s that utilized database procedures. Their proof-of-concept contributed to identifying duplicated code, improving system maintainability. For future directions, the author recommended using machine learning algorithms to fully automate the process and blackbox/whitebox testing methods to verify and validate the extracted microservices.

Al-Debagy et al. [34] decomposed monolithic into microservices architecture using a neural network model (code2vec). Their findings showed better results compared to other algorithms. Besides, authors validate their results by using quality measuring such as message level (CHM) and cohesion at domain level (CHD). Future directions may include further development and testing of the proposed model with other programming languages, as well as training.

Jin et al. [35] proposed the Functionality-oriented Service Candidate Identification (FoSCI) method, which uses a search-based functional atom grouping technique to identify service candidates from a monolithic system's execution traces. The authors assessed their method with an 8-metric service evaluation suite to analyze functionality, modularity, and evolvability. Additionally, the authors evaluated FoSCI against other methods (LIMBO, WCA, and MEM) to evaluate the impact of execution trace coverage on performance. Their results indicate that FoSCI outperforms the compared methods. However, their method prioritizes functionality over other quality attributes such as performance, security, and reliability, which could potentially benefit from future expansion.

Desai et al. [36] introduced a Graph Neural Network method for refactoring monolithic systems, termed COGCN (Clustering and Outlier-Aware Graph Convolution Net). This COGCN combines node representation, outlier detection, and clustering into a cohesive framework. To judge the quality of the clusters, four structural and desired metrics: modularity, structural modularity, non-extreme distribution (NED), and interface number (IFN) were used. Their results showed that COGCN works better than other methods like GCN, Node2Vector, and DeepWalk by improving the quality of clusters and finding outliers. Future work will include automatically determining microservice numbers and addressing procedural programming languages.

Kalia et al. [37] developed an approach called Mono2Micro. It used well-defined business use cases, spatio-temporal decomposition, and run-time call relations to functionally separate application classes. This method works better than other methods in well-defined metrics that are specific to the domain, such as FoSCI, MEM, CoGCN, and Bunch [17], [35], [36], [38] . Their tool was evaluated by using multiple criteria to verify its efficiency. Future directions focus on elaborating on the quality criteria, providing further assistance to develop effective use cases for practitioners, and generalizing the findings by supporting different programming languages.

Francisco et al. [39] performed a review of 20 papers addressing the migration process to microservices. Their results indicated that the majority of solutions rely on design

aspects, system dependency graphs, and clustering algorithms. Moreover, the majority of the studies depend on graphs and categorize them into microservices; 70% focus on web-based systems, mostly using Java as the main programming language. However, their study faces constraints due to its reliance on a single search engine and individual researcher bias. Furthermore, it advocates for a wider range of migration methodologies and database migration investigations.

Justas et al. [40] concentrated on the obstacles and approaches for migrating software to microservices by reviewing and analyzing different migration techniques; however, they didn't mention any particular evaluation metrics used to assess their review. Their results emphasized that the migration process is complex and expensive, with no one-size-fits-all solution. Future research includes a focus on developing standardized migration techniques to address the various issues posed by legacy systems.

Ren et al. [41] integrated static and dynamic analysis to examine the characteristics of monolithic systems. They employed function clustering to facilitate migration and hierarchical clustering to identify microservice candidates. They tested their method by comparing performance across four benchmark applications and validating the migration algorithm with 12 industrial and open-source applications. Their findings showed that their proposed method effectively migrates monolithic applications to microservices with high accuracy and low performance cost. However, the incomplete static analysis due to missing function invocations and user interactions may still limit the completeness of the migration.

Fritzsch et al. [42] conducted a qualitative study using semi-structured interviews in the context of migration to microservices. The authors conducted 16 comprehensive interviews with specialists from 10 software companies across 14 migration cases. Their finding indicated that the biggest motivations for migration were maintainability and scalability. Furthermore, many organizations choose a complete rebuild instead of a codebase breakdown. Principal problems included identifying the appropriate service cut and developing proficiency in emerging technologies. However, the sample procedure, which focuses only on 14 instances and individuals located in Germany, constrains their research and affects its generalizability.

Kalske et al. [43] performed a literature review focusing on architectural migration and associated challenges. Their findings indicate that organizations use microservices to mitigate complexity, enhance scalability, and resolve code ownership challenges. Nonetheless, restructuring and decoupling the tightly coupled monolith remain a significant challenge. However, their review lacks credibility and verification of their results, and there were no guidelines for conducting the review. Their future directions include understanding how various organizational structures influence the effectiveness of microservice adoption.

Vainio et al. [44] combined a case study and literature review to extract functionality from a monolithic system, then used microservices to demonstrate realistic real-world benefits and challenges for the migration process. Their findings showed significant advantages in scalability, maintainability, and the ability to use several programming languages for different services. Furthermore, the independent deployment of microservices improved system stability and performance. However, their case study did not thoroughly examine the architectural complexity and security implications of managing several microservices in extensive systems, indicating a need for additional research on these topics.

Eski et al. [45] employed a graph-based clustering methodology utilizing both static and evolutionary code coupling to derive microservices from monolithic applications. They evaluated their method by assessing two projects, which revealed an 89% success rate; however, certain services necessitated manual sub-clustering owing to their size. They planned to continue their research by adding weighted graph edges to their method, automating sub-clustering thresholds, and testing it on more projects using different algorithms.

Su et al. [46] conducted a systematic literature review by examining 32 studies to obtain insights into how software companies migrated from microservices back to monolithic architectures. Their study identified cost, complexity, scalability, performance, and organizational factors as the primary reasons for reverting. Future directions entail investigating the phenomenon across various industries and conducting empirical assessments to generalize the results.

Bandara et al. [47] developed a toolkit that utilized a fitness iterative function to identify microservices within monolithic systems. Their toolkit employed service quality metrics, including functionality, composability, self-containment, and usage. They compared the tool to manual microservice identification. Their results showed that their tool produced microservices like manual identification. Future directions include support for more programming languages and architectural patterns, machine learning models, and context knowledge to improve microservice extraction.

Michael Ayas et al. [48] examined the migration to microservices through interviews with 19 individuals from 16 organizations. Their results showed that the migration process is iterative and takes place on two levels: architectural and system-level. Moreover, they categorize key activities into four phases: designing architecture, altering the system, setting up supporting artifacts, and implementing technical artifacts. The study acknowledges the researchers' bias and the sample's representativeness. Future directions include examining more migration paths and validating results across organizations.

Taibi et al. [49] introduced a process-mining framework to aid in the decomposition of monolithic systems. Their framework identifies business processes based on log traces, clusters similar processes, and uses metrics to evaluate decomposition quality. A software architect validated their framework by comparing their results to a manual decomposition. Their framework revealed decomposition alternatives and software issues that manual analysis missed in an industrial case study. Future directions include automating the process, validating the methodology, and integrating patterns for microservices connectivity.

Silva et al. [50] conducted two case studies. These studies identified migration steps and challenges from legacy to

microservices. Their findings revealed four main issues: 1) identifying functionalities in large modules; 2) defining optimal microservice boundaries; 3) choosing features to convert into microservices; and 4) analyzing candidate microservice granularity and cohesion. Future steps include refining and supporting practical guidelines for migrating to microservice architectures, as well as surveying industry practitioners to learn more about migration challenges.

Kholy et al. [51] introduced the managing database for microservices architecture (MDMA) framework. Their framework indicates that MDMA significantly reduces execution time and data transfer size compared to a centralized approach and exhibits superior performance as the volume of requests increases. Furthermore, it improved flexibility and resilience, reducing the impact of service failures and facilitating data transfer during service deployment. Subsequent research may entail evaluating the framework in different real-world environments and broadening its applicability to various microservice architectures.

Antunes et al. [52] investigated the migration of a real-world application to a micro-frontend architecture. Results showed that micro-frontends enhance flexibility, team scalability, and incremental migration. Nonetheless, they observed increased complexity in the management of dependencies, environments, debugging, and testing. Future studies should investigate other projects and implementation methodologies, with a focus on simplifying dependency management and integration testing.

Maria et al. [53] investigated the migration towards micro-frontends. Their study demonstrates how to successfully re-implement a single-page application (SPA) using micro-frontends architecture by adopting Webpack to bundle modules and Cypress for testing. Their findings indicated improvements in team collaboration, independent deployment, and performance. Further studies include managing dependencies, enhancing integration, and evaluating performance.

Fritzsch et al. [54] used SLR to classify 10 existing approaches based on decomposition techniques. Their findings identified a lack of universally applicable approaches with adequate tool support. Future research will combine static code analysis with runtime data, create decomposition quality metrics, and automate the migration process.

Lauretis [55] presented a strategy for migrating to microservice. The author highlighted potential benefits such as improved scalability, maintainability, and evolvability for companies. Their strategy consists of five steps, including function analysis, business identification, business analysis, assigning functionalities, and creating microservices. However, their strategy is still in its early stages. Future directions include automating migration process and testing the strategy on real systems to gather performance and statistical data.

Nunes et al. [56] used transactional contexts to propose a microservices migration strategy. They used static code analysis to identify domain entities and applied a clustering algorithm to group them. Comparison with expert decompositions yields promising results. Future work will enhance the method and broaden the results, while limiting its applicability to specific frameworks and tools.

Santos et al. [57] proposed a complexity metric to measure the transition to microservices based on four similarity measures that examine entity decomposition. These measures focus on read and write sets, access sequences, and the cost of relaxing transactional consistency. Their metric was evaluated using three monolithic systems. Their complexity metric successfully identified the most complex decompositions. Future directions include using a combination of dynamic and static data, as well as further experimentation with a wider variety of systems.

Ma et al. [58] proposed Microservices Identification using Analysis for Database Access (MIADA) to account for the importance of "Database Per Service" in microservices design. Their approach facilitates clustering service endpoints for microservice identification. Two service-oriented software projects (PlanApproval and CoCoME) evaluated their results, demonstrating that MIADA can successfully recommend service endpoint clusters for microservice. Future directions include generating results and enhancing MIADA.

Ghofrani et al. [59] categorized migration challenges into inertia, anxiety, and context from 17 semi-structured expert interviews. The most significant barrier was migration anxiety, followed by inertia and context. Furthermore, they provided suggestions to overcome these obstacles. Future directions include evaluating their solutions and creating quality metrics.

Ghofrani et al. [60] conducted a survey of industry experts to identify challenges in microservices architecture. They identified critical challenges, including lack of notations, methods, and frameworks for microservices design, as well as insufficient tools for selecting third-party artifacts. In addition, they prioritized security, response time, and performance over resilience and fault tolerance. Future work includes expanding the scope and suggesting solutions for these gaps.

Razzaq et al. [61] provided a systematic mapping study on the migration towards microservices. Their review emphasized key benefits such as: including independent deployment, scalability, and lightweight mechanisms after migration. Besides, identifies migration challenges and success factors that help guide migration strategies. Future directions include, identify effective solutions for these areas, providing more in-depth research for microservice in emerging technologies.

V. RESULTS

As shown in Fig. 4, this SLR analyzed 50 studies published between 2018 and 2024 that focused on the migration process from monolithic to microservice. Furthermore, Table IV provides an overview of the findings from this literature research, covering the following topics:

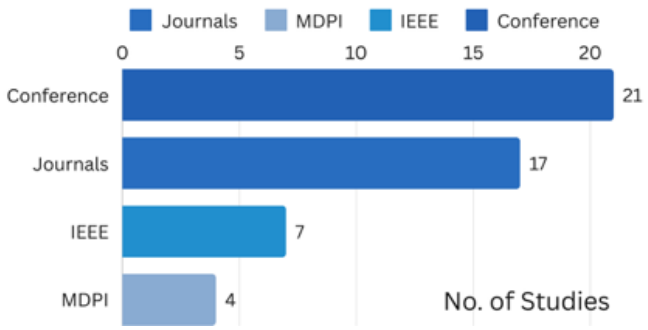


Fig. 4. Classification of studies examined in this SLR.

A. AI Algorithms and other Techniques:

Few authors emphasized providing proofs of concepts to support their ideas and provide insights for the migration process. Some studies conducted real-world industry case studies, attempting to assess and track the differences in performance, effort, scalability, and maintainability between monolithic and microservice architectures and drawing comparisons between them. Researchers are conducting extensive research on AI and other algorithmic techniques, including the indicator-based evolutionary algorithm (IBEA), neural networks, search-based algorithms, clustering algorithms, coupling, and cohesion. Additionally, some studies use SLR to improve their understanding of the microservice architecture. Fig. 5 presents the study distribution, demonstrating the use of AI and other clustering techniques in 18 studies, SLR in 13 studies, proof of concepts in 8 studies, and case studies in the remaining studies.

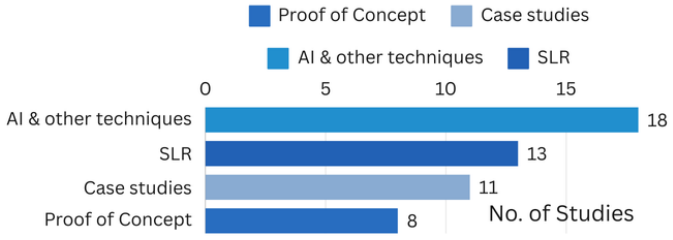


Fig. 5. Distribution of studies based on methodology.

B. Evaluation Techniques:

Major studies were evaluated and validated using a variety of techniques, including empirical assessment using a software system benchmark, the ISO/IEC 25012:2008 standard's definition of quality attributes, performance, scalability, and reliability comparisons, and architecture design comparisons. Other studies utilized cohesion at the domain level (CHD), conducted interviews with software industry experts, and utilized Postman tools for evaluation and testing. Fig. 6 presents the evaluation category.

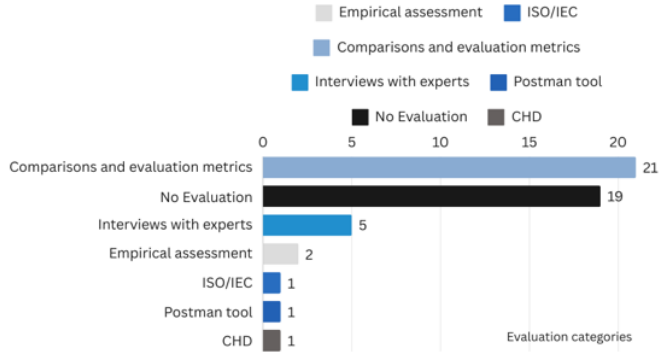


Fig. 6. Categorization of evaluations methods examined in this SLR.

C. Limitation and Future Work

Most studies in the software migration field, particularly those focusing on microservices, lack validation and verification of their models for real-world software systems, leading to a deficiency in generalizing the results of these studies. This gap emphasizes the need for further empirical research and actual implementations to test migration models in varied, real-world scenarios to ensure their robustness and reliability for wider use. The following is a list of future directions and areas for analysis and investigation:

1) Consider non-functional criteria.
2) Perspectives of software architects and developers.
3) Automated tools are needed for manual decomposition.
4) Generalizing results by testing on real-world industry.
5) The impact of cloud servers on the migration process is significant. Hyperscale's such as Azure, AWS, and GCP offer unique features like scalability options, flexible resource allocation, and Devops tools.
6) Adopt micro-frontend alongside microservices.

The subsequent discussion and answers correspond to the research questions outlined in Section III:

RQ1: There is a significant shortage of research on AI and ML methods in the software migration to microservices domain. Few studies are employing algorithms like the indicator-based evolutionary algorithm (IBEA), search-based techniques, and clustering algorithms. Other studies rely on comparisons and expert inputs to assess software quality criteria. The remaining studies offer proof-of-concepts but do not provide a comprehensive solution. A comprehensive solution would involve an automated method or algorithm that utilizes AI to simplify the transition from monolithic to microservice processes while maintaining performance, scalability, and availability. Additionally, real-world industry software systems must evaluate the solutions.

RQ2: The most significant challenges and obstacles in this domain arise from how software warehouses and architects

break down the logic and databases of software systems into dozens of interconnected microservices, which complicates the assurance of effective communication, data consistency, and the preservation of the system's overall integrity. Furthermore, it's crucial to consider the manner in which these services will interact and coordinate with each other. In addition, microservice architecture must meet non-functional criteria like performance, scalability, security, and reliability. As suggested in [7], AI can assist by using powerful ML techniques such as NLP to analyze the code, documentation, and system requirements, as well as complex clustering techniques to decompose the system logics into multiple services. To address non-functional criteria, it may be beneficial to monitor and utilize technologies, such as cloud computing and DevOps methodologies.

RQ3: The evaluation of software migration from monolithic to microservices is mostly based on comparing the two architectures and soliciting experts' opinions and recommendations. Despite this, AI models and other algorithms need to be rigorously evaluated and tested in the real industry to gather feedback from organizations and end users, enabling us to monitor performance and obtain relevant insights.

VI. Conclusion

This study conducted a systematic literature review to explore the key challenges, obstacles, and improvements in the software industry, specifically the migration from monolithic to microservice architecture. While there is growing interest in developing AI and other algorithmic techniques to facilitate this migration process, the field remains in its early stages, lacking comprehensive, end-to-end solutions that address the full complexity of the process. Most existing studies focus on theoretical frameworks, proof-of-concepts, and expert judgments rather than verifying and validating these approaches through real-world implementation.

Table IV provides a summary of this literature review, which includes a list of studies, each describing the methods used, the evaluation technologies, the results, their limitations, and future work. Section V addresses detailed answers to the three research questions. Both would assist software developers and other authors by providing the foundation for a comprehensive solution for addressing migration challenges.

The results of this literature study highlight a significant gap in the creation of effective migration tools and the validation of these solutions within the software industry. This emphasizes the need for future research to focus on the development, evaluation, and validation of automated tools in real-world environments. Furthermore, the migration process must emphasize non-functional requirements such as performance, scalability, reliability, and security. Integrating DevOps methodologies with cloud computing concepts is essential for optimizing the advantages of microservice architectures.

By identifying these gaps, this review emphasizes the importance of addressing both the theoretical and practical dimensions of the migration process, providing a roadmap for future research aimed at enhancing the efficiency, effectiveness, and scalability of software migrations. Resolving these challenges will ultimately enhance the industry's ability to migrate from monolithic systems to modern microservice-based architectures.

TABLE IV. A Concise Summary of the Research Findings from the Literature

R	Techniques used	Results	Limitations and future work
[1]	IBEA	Empirical testing showed the proposed approach is more effective.	Considering non-functional and software evaluation using.
[19]	Proof-of- concept	Migrate monolithic data storage to microservice can be applied.	Automation is needed, and the results are not generalizable.
[18]	Literature Review	Patterns for microservices. Microservices vs. monolithic pros/cons.	There is a need for evaluation, automatic decomposition.
[20]	Combined approaches	Customization-driven migration can guide monoliths into SaaS.	Generalizing results and automating the migration process.
[21]	Literature Review	List of challenges and benefits that arise in the migration process.	Analyze architectures pros and cons and automate migration.
[11]	Case Study	Modular monoliths aid migration, consistency, and communication.	Generalize and apply results to other case studies.
[12]	Case Study	Monolithic outperforms microservices, Java is better at computation.	Increase system complexity and deploy to various clouds.
[22]	Case Study	Their model optimizes development processes.	Generalizing the result required model improvement.
[8]	Case Study	Microservices improved scalability and solved the monolithic issues.	Expanding proof of concepts to generalize the results.
[10]	Search-based techniques	Promising strategy could be applied to other monolithic systems.	The model needs to be configurable and generalized.
[23]	Case Study	The migration affects stakeholders. Upload time affects deployment.	Cloud servers' application support and generalize the findings.
[24]	Case Study	Monolithic suits small-medium apps, but microservices are scalable.	Real-world testing. Ensuring security among microservices.
[25]	Literature Review	Microservices migration in its early stages, with few methods exists.	Consider deployment and microservice identification metrics.
[26]	Case Study	Microservices boost hardware efficiency, productivity, and cost.	Develop migration tools is needed and secure microservices.
[27]	Proof-of- concept	The conceptual model solve system breakdown migration steps.	Software business logic extraction tool need.

[28]	Decision framework	A complete microservices adoption measuring set for enterprises.	Validating framework and adding cloud-native technologies.
[7]	Clustering techniques	Microservices work better and stress business processes.	Using AI to identify microservices and security concerns.
[29]	Coupling and cohesion	The suggested metrics criteria are better for industrial use.	Evaluates microservice assessment using real-world case studies.
[30]	Literature Review	Emphasize automated identification and evaluation standards.	Studies focused on specific challenges and lack migration tools.
[31]	Literature Review	The study shows that categorization helps industrial professionals.	Grouping services by type may help generalize results.
[32]	Clustering techniques	Results showed the proposed model outperformed others.	Generalizing the results.
[33]	Proof-of- concept	The model found duplication and improved maintainability.	Automation using ML and black/white box for validation.
[34]	Neural network	The proposed algorithm performed better than others.	Testing the model in other languages and training on source codes.
[37]	Clustering techniques	Mono2Micro outperforms other methods.	Elaborate on the quality criteria and develop effective use cases.
[35]	Search-based techniques	FoSCI is better in service quality, functionality, and modularity.	Performance, security, and reliability are ignored.
[36]	Clustering techniques	COGCN improves cluster quality and outlier identification.	Determining microservice numbers and procedural languages.
[16]	Clustering techniques	FoME provides cohesive microservices with looser coupling.	Ensure the use of high-quality test cases and automate the process.
[39]	Literature Review	Most research used design, dependency graphs, and clustering.	Biased results and more migration methods are advocated.
[40]	Literature Review	Migration is complicated and expensive, with no single solution.	Standard migration methods to address legacy system issues.
[41]	Clustering techniques	Their microservice migration method was accurate with low cost.	Generalizing the results by adopting user interactions.
[42]	Qualitative study	Maintenance and scalability encouraged migration.	Generalizability is limited by the 14-person German sample.
[43]	Literature Review	For simplicity, scalability, and ownership, firms use microservices.	Their results lack credibility and verification.
[44]	Case Study	Microservices are scalable, maintainable, stable, and multilingual.	Security and architectural complexity need more research.
[45]	Clustering techniques	Their method revealed 89% success rate of extracted microservice.	Testing on different clustering algorithms to enhance the model.
[46]	Literature Review	Cost, complexity, and organization drive monolithic Reverting.	Examining the phenomenon across industries and generalizing.
[47]	Toolkit	Toolkit produced microservices like manual identification.	Supporting more languages, patterns, ML models, and context.
[48]	Qualitative study	Migration occurs on two levels: architectural and system-level.	Researcher bias and sample representativeness affect findings.
[49]	Process-mining	Framework outperforms manual analysis in industrial case studies.	Automating the process, validating, and integrating more patterns.
[50]	Case Study	Module's size, boundaries, features, and cohesion are challenges.	Survey industry specialists and refine microservice migration.
[51]	MDMA framework	MDMA improve runtime, data transfer, flexibility.	Testing the framework in real-world and expanding its use.
[52]	Case Study	Micro-frontends improve migration, scalability, and flexibility.	Dependencies, debugging, and testing require further research.
[53]	Case Study	SPA enhances deployment, collaboration, and performance.	Managing dependencies, enhancing integration and performance.
[54]	Literature Review	Lack of universally approaches with adequate tool support.	Automate migration with static code analysis and runtime data.
[55]	Proof-of- concept	Improved Company scalability, maintainability, and evolution.	Automating migration process and testing on real systems.
[56]	Clustering techniques	Comparison with expert decompositions yields promising results.	Improve the method and generalize the results.
[57]	Clustering techniques	Their metric successfully identified the complex decompositions.	Using dynamic and static data, along with further investigation.
[58]	Clustering techniques	MIADA can successfully recommend service endpoint clusters.	Future directions include generating results and enhancing MIADA.
[59]	Qualitative study	Anxiety, inertia, and context were the most barriers in migration.	Evaluating their proposed solutions and creating quality metrics.
[60]	Literature Review	Microservices design notations and method were critical challenges.	Expanding the scope and suggesting solutions for these gaps.
[61]	Literature Review	Emphasized independent deployment, scalability, and lightweight.	Provide research in emerging technologies to find solutions.

REFERENCES

[1] K. Sellami, A. Ouni, M. A. Saied, S. Bouktif, and M. W. Mkaouer, "Improving microservices extraction using evolutionary search," Inf Softw Technol, vol. 151, Nov. 2022, doi: 10.1016/j.infsof.2022.106996.

[2] S. Newman, "Building Microservices SECOND EDITION Designing Fine-Grained Systems."

[3] F. Ponce, J. Soldani, H. Astudillo, and A. Brogi, "Smells and Refactorings for Microservices Security: A Multivocal Literature Review," Apr. 2021, [Online]. Available: http://arxiv.org/abs/2104.13303

[4] M. G. de Almeida and E. D. Canedo, "Authentication and Authorization in Microservices Architecture: A Systematic Literature Review," Applied Sciences (Switzerland), vol. 12, no. 6, Mar. 2022, doi: 10.3390/app12063023.

[5] I. Saidani, A. Ouni, M. W. Mkaouer, and A. Saied, "Towards Automated Microservices Extraction Using Muti-objective Evolutionary Search," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer, 2019, pp. 58–63. doi: 10.1007/978-3-030-33702-5_5.

[6] S. Li et al., "A dataflow-driven approach to identifying microservices from monolithic applications," Journal of Systems and Software, vol. 157, Nov. 2019, doi: 10.1016/j.jss.2019.07.008.

[7] M. Daoud, A. El Mezouari, N. Faci, D. Benslimane, Z. Maamar, and A. El Fazziki, "A multi-model based microservices identification approach," Journal of Systems Architecture, vol. 118, Sep. 2021, doi: 10.1016/j.sysarc.2021.102200.

[8] M. Mazzara, N. Dragoni, A. Bucchiarone, A. Giaretta, S. T. Larsen, and S. Dustdar, "Microservices: Migration of a Mission Critical System," IEEE Trans Serv Comput, vol. 14, no. 5, pp. 1464–1477, 2021, doi: 10.1109/TSC.2018.2889087.

[9] P. Di Francesco, P. Lago, and I. Malavolta, "Migrating Towards Microservice Architectures: An Industrial Survey," in Proceedings - 2018 IEEE 15th International Conference on Software Architecture, ICSA 2018, Institute of Electrical and Electronics Engineers Inc., Jul. 2018, pp. 29–38. doi: 10.1109/ICSA.2018.00012.

[10] W. K. G. Assunção et al., "A Multi-Criteria Strategy for Redesigning Legacy Features as Microservices: An Industrial Case Study," in Proceedings - 2021 IEEE International Conference on Software Analysis, Evolution and Reengineering, SANER 2021, Institute of Electrical and Electronics Engineers Inc., Mar. 2021, pp. 377–387. doi: 10.1109/SANER50967.2021.00042.

[11] D. Faustino, N. Gonçalves, M. Portela, and A. Rito Silva, "Stepwise migration of a monolith to a microservice architecture: Performance and migration effort evaluation," Performance Evaluation, vol. 164, May 2024, doi: 10.1016/j.peva.2024.102411.

[12] G. Blinowski, A. Ojdowska, and A. Przybylek, "Monolithic vs. Microservice Architecture: A Performance and Scalability Evaluation," IEEE Access, vol. 10, pp. 20357–20374, 2022, doi: 10.1109/ACCESS.2022.3152803.

[13] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering - A systematic literature review," Jan. 2009. doi: 10.1016/j.infsof.2008.09.009.

[14] Y. Xiao and M. Watson, "Guidance on Conducting a Systematic Literature Review," Mar. 01, 2019, SAGE Publications Inc. doi: 10.1177/0739456X17723971.

[15] Z. Bai and T. Wasson, "Conducting Systematic Literature Reviews in Information Systems: An Analysis of Guidelines," 2019. [Online]. Available: https://www.researchgate.net/publication/340686721

[16] W. Jin, T. Liu, Q. Zheng, D. Cui, and Y. Cai, "Functionality-Oriented Microservice Extraction Based on Execution Trace Clustering," in Proceedings - 2018 IEEE International Conference on Web Services, ICWS 2018 - Part of the 2018 IEEE World Congress on Services, Institute of Electrical and Electronics Engineers Inc., Sep. 2018, pp. 211–218. doi: 10.1109/ICWS.2018.00034.

[17] G. Mazlami, J. Cito, and P. Leitner, "Extraction of Microservices from Monolithic Software Architectures," in Proceedings - 2017 IEEE 24th International Conference on Web Services, ICWS 2017, Institute of Electrical and Electronics Engineers Inc., Sep. 2017, pp. 524–531. doi: 10.1109/ICWS.2017.61.

[18] V. Velepucha and P. Flores, "A Survey on Microservices Architecture: Principles, Patterns and Migration Challenges," IEEE Access, vol. 11, pp. 88339–88358, 2023, doi: 10.1109/ACCESS.2023.3305687.

[19] J. Kazanavičius, D. Mažeika, and D. Kalibatienė, "An Approach to Migrate a Monolith Database into Multi-Model Polyglot Persistence Based on Microservice Architecture: A Case Study for Mainframe Database," Applied Sciences (Switzerland), vol. 12, no. 12, Jun. 2022, doi: 10.3390/app12126189.

[20] E. T. Nordli, S. G. Haugeland, P. H. Nguyen, H. Song, and F. Chauvel, "Migrating monoliths to cloud-native microservices for customizable SaaS," Inf Softw Technol, vol. 160, Aug. 2023, doi: 10.1016/j.infsof.2023.107230.

[21] V. Velepucha and P. Flores, "Monoliths to microservices-Migration Problems and Challenges: A SMS," in Proceedings - 2021 2nd International Conference on Information Systems and Software Technologies, ICI2ST 2021, Institute of Electrical and Electronics Engineers Inc., Mar. 2021, pp. 135–142. doi: 10.1109/ICI2ST51859.2021.00027.

[22] A. Bastidas Fuertes, M. Pérez, and J. Meza, "Transpiler-Based Architecture Design Model for Back-End Layers in Software Development," Applied Sciences (Switzerland), vol. 13, no. 20, Oct. 2023, doi: 10.3390/app132011371.

[23] Teguh Prasandy, Titan, Dina Fitria Murad, and Taufik Darwis, "Migrating Application from Monolith to Microservices," in 2020 International Conference on Information Management and Technology (ICIMTech), Bandung, Indonesia: IEEE, 2020, pp. 726–731. doi: 10.1109/ICIMTech50083.2020.9211252.

[24] A. B. Fondazione et al., "Migration from Monolith to Microservices : Benchmarking a Case Study", doi: 10.13140/RG.2.2.27715.14883.

[25] Y. Abgaz et al., "Decomposition of Monolith Applications Into Microservices Architectures: A Systematic Review," IEEE Transactions on Software Engineering, vol. 49, no. 8, pp. 4213–4242, Aug. 2023, doi: 10.1109/TSE.2023.3287297.

[26] F. Tapia, M. ángel Mora, W. Fuertes, H. Aules, E. Flores, and T. Toulkeridis, "From monolithic systems to microservices: A comparative study of performance," Applied Sciences (Switzerland), vol. 10, no. 17, Sep. 2020, doi: 10.3390/app10175797.

[27] D. Kuryazov, D. Jabborov, and B. Khujamuratov, "Towards Decomposing Monolithic Applications into Microservices," in 14th IEEE International Conference on Application of Information and Communication Technologies, AICT 2020 - Proceedings, Institute of Electrical and Electronics Engineers Inc., Oct. 2020. doi: 10.1109/AICT50176.2020.9368571.

[28] F. Auer, V. Lenarduzzi, M. Felderer, and D. Taibi, "From monolithic systems to Microservices: An assessment framework," Inf Softw Technol, vol. 137, Sep. 2021, doi: 10.1016/j.infsof.2021.106600.

[29] M. H. Hasan, M. Hafeez Osman, N. I. Admodisastro, and S. Muhammad, "From Monolith to Microservice: Measuring Architecture Maintainability," 2023. [Online]. Available: www.ijacsa.thesai.org

[30] I. Oumoussa and R. Saidi, "Evolution of Microservices Identification in Monolith Decomposition: A Systematic Review," IEEE Access, vol. 12, pp. 23389–23405, 2024, doi: 10.1109/ACCESS.2024.3365079.

[31] M. Abdellatif et al., "A taxonomy of service identification approaches for legacy software systems modernization," Journal of Systems and Software, vol. 173, Mar. 2021, doi: 10.1016/j.jss.2020.110868.

[32] J. Li, H. Xu, X. Xu, and Z. Wang, "A Novel Method for Identifying Microservices by Considering Quality Expectations and Deployment Constraints," https://doi.org/10.1142/S021819402250019X, vol. 32, no. 3, pp. 417–437, Apr. 2022, doi: 10.1142/S021819402250019X.

[33] M. H. Gomes Barbosa and P. H. M. Maia, "Towards Identifying Microservice Candidates from Business Rules Implemented in Stored Procedures," in Proceedings - 2020 IEEE International Conference on Software Architecture Companion, ICSA-C 2020, Institute of Electrical and Electronics Engineers Inc., Mar. 2020, pp. 41–48. doi: 10.1109/ICSA-C50368.2020.00015.

[34] O. Al-Debagy and P. Martinek, "A Microservice Decomposition Method Through Using Distributed Representation Of Source Code," Scalable Computing, vol. 22, no. 1, pp. 39–52, 2021, doi: 10.12694:/scpe.v22i1.1836.

[35] W. Jin, T. Liu, Y. Cai, R. Kazman, R. Mo, and Q. Zheng, "Service Candidate Identification from Monolithic Systems Based on Execution Traces," IEEE Transactions on Software Engineering, vol. 47, no. 5, pp. 987–1007, May 2021, doi: 10.1109/TSE.2019.2910531.

[36] U. Desai, S. Bandyopadhyay, and S. Tamilselvam, "Graph Neural Network to Dilute Outliers for Refactoring Monolith Application," Feb. 2021, [Online]. Available: http://arxiv.org/abs/2102.03827

[37] A. K. Kalia, J. Xiao, R. Krishna, S. Sinha, M. Vukovic, and D. Banerjee, "Mono2Micro: A practical and effective tool for decomposing monolithic Java applications to microservices," in ESEC/FSE 2021 - Proceedings of the 29th ACM Joint Meeting European Software Engineering Conference and Symposium on the Foundations of Software Engineering, Association for Computing Machinery, Inc, Aug. 2021, pp. 1214–1224. doi: 10.1145/3468264.3473915.

[38] B. S. Mitchell and S. Mancoridis, "On the Automatic Modularization of Software Systems Using the Bunch Tool."

[39] P. Francisco, M. Gastón, and A. Hernán, Migrating from monolithic architecture to microservices: A Rapid Review. IEEE, 2019. doi: 10.1109/SCCC49216.2019.8966423.

[40] K. Justas and M. Dalius, "Migrating Legacy Software to Microservices Architecture," Institute of Electrical and Electronics Engineers, 2019, p. 32. doi: 10.1109/eStream.2019.8732170.

[41] Z. Ren et al., "Migrating web applications from monolithic structure to microservices architecture," in ACM International Conference Proceeding Series, Association for Computing Machinery, Sep. 2018. doi: 10.1145/3275219.3275230.

[42] J. Fritzsch, J. Bogner, S. Wagner, and A. Zimmermann, "Microservices Migration in Industry: Intentions, Strategies, and Challenges," in Proceedings - 2019 IEEE International Conference on Software Maintenance and Evolution, ICSME 2019, Institute of Electrical and Electronics Engineers Inc., Sep. 2019, pp. 481–490. doi: 10.1109/ICSME.2019.00081.

[43] M. Kalske, N. Mäkitalo, and T. Mikkonen, "Challenges When Moving from Monolith to Microservice Architecture," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, 2018, pp. 32–47. doi: 10.1007/978-3-319-74433-9_3.

[44] M. Vainio and T. Antti-Pekka, "The benefits and challenges in migrating from a monolithic architecture into microservice architecture," 2021. [Online]. Available: http://www.cs.helsinki.fi/

[45] S. Eski and F. Buzluca, "An automatic extraction approach - Transition to microservices architecture from monolithic application," in ACM International Conference Proceeding Series, Association for Computing Machinery, 2018. doi: 10.1145/3234152.3234195.

[46] R. Su, X. Li, and D. Taibi, "From Microservice to Monolith: A Multivocal Literature Review †," Apr. 01, 2024, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/electronics13081452.

[47] C. Bandara and I. Perera, "Transforming monolithic systems to microservices - An analysis toolkit for legacy code evaluation," in 20th International Conference on Advances in ICT for Emerging Regions, ICTer 2020 - Proceedings, Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 95–100. doi: 10.1109/ICTer51097.2020.9325443.

[48] H. Michael Ayas, P. Leitner, and R. Hebig, "The Migration Journey Towards Microservices," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Science and Business Media Deutschland GmbH, 2021, pp. 20–35. doi: 10.1007/978-3-030-91452-3_2.

[49] D. Taibi and K. Systä, "From monolithic systems to microservices: A decomposition framework based on process mining," in CLOSER 2019 - Proceedings of the 9th International Conference on Cloud Computing and Services Science, SciTePress, 2019, pp. 153–164. doi: 10.5220/0007755901530164.

[50] H. H. O. S. Da Silva, G. F. De Carneiro, and M. P. Monteiro, "Towards a roadmap for the migration of legacy software systems to a microservice based architecture," in CLOSER 2019 - Proceedings of the 9th International Conference on Cloud Computing and Services Science, SciTePress, 2019, pp. 37–47. doi: 10.5220/0007618400370047.

[51] M. El Kholy and A. El Fatatry, "Framework for Interaction between Databases and Microservice Architecture," IT Prof, vol. 21, no. 5, pp. 57–63, Sep. 2019, doi: 10.1109/MITP.2018.2889268.

[52] F. Antunes, M. J. Dias De Lima, M. Antônio, P. Araújo, D. Taibi, and M. Kalinowski, "Investigating Benefits and Limitations of Migrating to a Micro-Frontends Architecture," 2024.

[53] A. Maria and C. Fulvio, "Exploring Software Architectural Transitions: From Monolithic Applications to Microfrontends enhanced by Webpack library and Cypress Testing," Jul. 2024.

[54] J. Fritzsch, J. Bogner, A. Zimmermann, and S. Wagner, "From Monolith to Microservices: A Classification of Refactoring Approaches," 2018. doi: arXiv:1807.10059.

[55] L. De Lauretis, "From monolithic architecture to microservices architecture," in Proceedings - 2019 IEEE 30th International Symposium Software Reliability Engineering Workshops, ISSREW 2019, Institute of Electrical and Electronics Engineers Inc., Oct. 2019, pp. 93–96. doi: 10.1109/ISSREW.2019.00050.

[56] L. Nunes, N. Santos, and A. Rito Silva, "From a monolith to a microservices architecture: An approach based on transactional contexts," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019. doi: 10.1007/978-3-030-29983-5_3.

[57] N. Santos and A. Rito Silva, "A complexity metric for microservices architecture migration," in Proceedings - IEEE 17th International Conference on Software Architecture, ICSA 2020, Institute of Electrical and Electronics Engineers Inc., Mar. 2020, pp. 169–178. doi: 10.1109/ICSA47634.2020.00024.

[58] S. P. Ma, T. W. Lu, and C. C. Li, "Migrating Monoliths to Microservices based on the Analysis of Database Access Requests," in Proceedings - 16th IEEE International Conference on Service-Oriented System Engineering, SOSE 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 11–18. doi: 10.1109/SOSE55356.2022.00008.

[59] J. Ghofrani and A. Bozorgmehr, "Migration to Microservices: Barriers and Solutions," 2019, pp. 269–281. doi: 10.1007/978-3-030-32475-9_20.

[60] J. Ghofrani and D. Lübke, "Challenges of Microservices Architecture: A Survey on the State of the Practice," 2018. [Online]. Available: http://ceur-ws.org/Vol-2072

[61] A. Razzaq and S. A. K. Ghayyur, "A systematic mapping study: The new age of software architecture from monolithic to microservice architecture—awareness and challenges," Mar. 01, 2023, John Wiley and Sons Inc. doi: 10.1002/cae.22586.

A Review on NS Beyond 5G: Techniques, Applications, Challenges and Future Research Directions

Cui Zhiyi, Azana Hafizah Mohd Aman, Faizan Qamar*
Center for Cyber Security-Faculty of Information Science and Technology (FTSM),
Universiti Kebangsaan Malaysia (UKM), 43600 UKM Bangi, Selangor, Malaysia

***Abstract*—With the advent of the fifth generation (5G) era, many emerging Internet of Things (IoT) applications have emerged to make life more convenient and intelligent. While the number of connected devices is growing, there are also differences in the network requirements of each device. Network slicing (NS) as an emerging technology, provides multiple logical networks for infrastructure. Each of these logical networks can provide specialized services for different needs of different applications by defining its logical topology, reliability and security level. This article provides an overview of the basic architecture, categories, and life cycle of network slicing. Then summarized two kinds of resource allocation methods and security problems based on three kinds of network slicing technologies. With the investigation of recent studies, it is found that network slicing is widely used in Industrial Internet of Things (IIoT), Internet of Medical Things (IoMT), in-vehicle systems and other applications. It improves network efficiency, improves service quality and enhances security and privacy by optimizing indicators such as latency, resource management and service quality. At the end of this study, according to the challenges faced by different research methods, the future research direction of network slicing is proposed.**

***Keywords*—5G; network slicing; resource allocation; dynamic allocation; security**

I. INTRODUCTION

5G and higher networks are predicted to have high throughput, low latency, and the power to concatenate many devices [1]. With the popularization of 5G, more and more new services appear in our lives. These different services put forward a variety of requirements for the quality of service (QoS) of the network [2]. For healthcare, autonomous industry, and smart vehicle services, networks need to offer lower latency and high reliability, security, and privacy [3]. Smart city and grid services have many applications, requiring the network to be highly scalable and provide large data sets [4]. Smart home needs safe short-distance service, and smart weather station needs long-distance communication with low bandwidth [5]. Therefore, 5G networks need to establish a more flexible and guaranteed dynamic network to meet different needs of all walks of life [6].

In this context, NS technology comes into being. NS can divide the physical network into multiple logical networks, and each logical network can formulate its own functions and characteristics to provide an independent network for a specific application [7]. However, in the highly fragmented network environment, vast demand for specific traffic leads to overload, which affects the quality of service of the network [8]. Therefore, many NS researches achieve real-time dynamic resource allocation and prediction by using artificial intelligence (AI) or machine language (ML) algorithms [9]. The algorithm predicts traffic patterns and historical data based on future demand, thereby adjusting resources ahead of time to ensure that the slice maintains stable performance during peak periods [10]. The creation of different slices, configuration updates, etc. become more complex as the number and types of slices increase. AI-driven network orchestration and autonomous management tools to adjust the lifecycle of slices, and propose automated orchestration systems and intelligent management platforms to reduce the complexity of human intervention and management [11].

In addition, NS has compatibility problems with existing network infrastructure, and operators use virtualization and software-defined networking (SDN) technologies to transfer traditional hardware functions to programmable virtual platforms to achieve flexible network control and rapid deployment [12]. This strategy not only improves the adaptability and security of the network, but also lays a technical foundation for the realization of large-scale NS in the future.

This paper introduces the architecture, function and practical application of NS. Summarizes the function and architecture of NS by reviewing the progress of research on NS in recent years. In this paper, reviews the practical applications of NS, pointes out the challenges of applying NS, and prospects the future research directions. On this basis, the major contributions of this paper are as follows:

- Introduces the architecture and life cycle of NS and classifies it according to different criteria.
- Investigate two resource allocation methods based on NS, introduce the methods in detail and summarize the test results.
- Surveys and summarizes how NS can be applied to healthcare, industrial application and in-vehicle system services.

*Corresponding Author

- Introduces the security challenges of basic functions of NS such as SDN, network functions virtualization (NFV) and their hybrid models.

This paper summarizes the development of NS and the research projects and achievements based on different methods. The remainder of this paper is as follows. Section II surveys the existing literature on NS and compares the classification from the technical point of view. Section III introduces the 5G NS architecture, classification based on different criteria, life cycle and introduces two resource allocation problems based on NS. Section IV introduces some technology based on NS. Section V introduces the security issues under the three techniques. Section VI further discusses the prior research. Section VII proposes the future research directions and challenges. Section VIII summarizes the paper.

II. Related Work

In [13], summarizes the basic concepts and analyzes implementation of NS. According to the standard given by NGMN, the network slice is divided into the infrastructure layer, network slice instance layer and service instance layer. Two categories of functional NS, Radio Access Network (RAN) subslices and core subslices, are also introduced. The authors summarize the four life cycles of NS: preparation, debugging, operation and decommissioning. The last part analyzes the application of NS in IoT and other emerging technologies such as blockchain, AI, and ML, and finds the challenges faced by each part and the direction that can be tried in the survey.

In [14], makes a summary of the six advantages of NS and their problems. The dynamic allocation of free resources among slices optimizes the scalability of the network. Use ETSI's zero-touch Network and Service Management (ZSM) to reduce human intervention and automate network operations to address scalability challenges. Or use a distributed Artificial Intelligence (DAI) approach to solve orchestration requests in the NS manager. Each slice has different network functions (NFS) and attributes to satisfy the different requirements of different applications. The reinforcement learning algorithm is used to identify slice attributes and recursion to achieve this goal. The isolation mechanism of dedicated slices can mitigate IoT security attacks. Research on design security and operational security will be expanded in the future. At the same time, it uses the verification function of the slice itself, Bayesian network, and autoencoder for intrusion detection and anomaly development, as in the case of the Application Programming Interface (API) used by the third-party tenant. Dedicated slices are assigned to each application, and isolation mechanisms and communication schemes are optimized to protect privacy. Zero-knowledge proof and argumentation tools for knowledge can optimize inter-slice communication. NS assigns each application a specific slice with sufficient resources to guarantee QoS.

In [15], the research on solving NS management in recent years is summarized. Due to the high-speed and multi-connection characteristics of 5G networks, as well as the flexibility and cost of its architecture. The authors propose that one of the challenges of NS is to completely isolate NS, in which each slice instance has its own function and is not shared between slices. But at the end of the survey, the authors put forward many shortcomings about NS. For example, the resource allocation fast mobile service cannot meet the needs of all the slicing networks, so that the compatibility of this service has drawbacks. 5G does not support inter-slice communication when performing isolation, and information propagation will be subjected to security attacks. Limited end-to-end slice orchestration also becomes an obstacle to solve this challenge simultaneously. Future work attempts to use network simulation to analyze the 5G NS architecture in different mobility scenarios, divide the 5G network into a large number of slices, and let these slices operate independently.

In [16], the scarce overview of NS's security aspects is summarized, and new security and privacy challenges based on NS are introduced for the first time. The investigation points out that when NS processes a large number of data sets because tenants use public resources, it will lead to information leakage and cause privacy security issues. The authors propose a security taxonomy based on NS, Categories include NS attack scenarios security threats, challenges and issues security solutions and trust and privacy in NS. At the same time, attack scenarios such as location tracking, impersonation attack, traditional interworking, and sealing between multiple slices are introduced. Finally, the security solutions based on AI/ML are summarized. Examples include domain isolation against DoS, authentication against identity spoofing, and access control for intrusion detection.

In [17], the author introduces some ML methods about wireless access network slice resource management. Evolutionary algorithms, supervised and unsupervised learning, reinforcement and distributed learning are used to solve the problems of slice resource sharing, dynamic management and isolation security in 5G/B5G. These ML methods are dynamic online techniques which can interact with dynamic network changes and select the most effective decision. To guarantee the QoS, the resources of a slice should be kept constant during a certain time to avoid its traffic load changes affecting other slices. Using Long Short-Term Memory (LSTM) to predict the user's demand for resources to reserve resources can reduce the number of resource reconfiguration. Many projects fixes the transmission power of user equipment, and the dynamic power allocation method combines the channel environment and interference conditions through iterative and online algorithms to optimize the drop situation and anti-interference ability. The user mobility management of slices can be optimized by predicting the user mobility to optimize the resource allocation rate and guarantee QoS. Priority queues prioritize slices to ensure service quality.

In [18], it is pointed out that the attacker attacks the network by using the NS function, which leads to the system failure. The authors divide the attacks on NS into inter-frame attacks, intra-frame attacks and lifetime attacks. At the same time, the solutions are divided into three types: RAN, core network and general solution. At this stage, NS security is mainly maintained using resource isolation, ML and encryption algorithm. This paper evaluates the performance of solving malicious attacks using open interfaces. The experimental results show that the RAN with performance isolation only needs half of the resources of the RAN without performance

isolation to achieve the same throughput. The incoming requests of mobile users were processed by first in First Out (FIFO) queue, and the statistical distribution and isolation forest were used to detect outliers. The results show that isolation forest produces more outliers. NS's end-to-end isolation can achieve security in complex systems, and blockchain can prevent data breaches and automatically manage Service Level agreements (SLAs). The existing cryptosystems cannot well protect NS from security intrusion and using quantum secure encryption method as an alternative is a future research goal.

In [19], the study explored the implementation of cloud-native NS based on SDN and multi-access Edge computing (MEC). These two approaches enable NS by virtualizing and dynamically configuring network resources. The authors discuss the design principles of cloud-native 5G core networks, how to optimize network function modularity and service chain through microservice architecture simultaneously. This paper proposes an SDN-based MEC architecture to realize low-latency and Ultra-reliable Delay Communication (URLLC) services. The feasibility of the proposed architecture in terms of NS mobility and service migration is verified through an initial evaluation in multiple edge cloud deployments in Southeast Australia. Cloud-native NS has four main challenges: federated slicing, dynamic service chaining, edge computing, and federated resource management. To address these challenges, the main goals of future research on NS are to enhance the slice federalization mechanism, optimize the dynamic service chain management, promote edge intelligent computing, and realize 4C resource joint management. Blockchain and AI technology are used to enhance resource management and isolation mechanisms. Develop more efficient dynamic service chain management algorithms in 5G architecture, and design new neural network and ML algorithms to meet low latency, high computing requirements and the management of communication, computing, caching and control resources in the joint edge environment, so as to optimize resource utilization. A summary of each literature is shown in Table I.

It is found that different methods have different advantages and disadvantages in terms of effectiveness, efficiency, scalability and applicability. For example, literature [13] focuses on the basic concept and implementation of NS, which, despite its high effectiveness, lacks in terms of scalability. The reinforcement learning method proposed in literature [14] has excellent performance in improving efficiency and scalability, but there are still security challenges. Literature in [15] presents the high efficiency and applicability of independent slices under the high connection characteristics of 5G networks. Nevertheless, the inadequate resource allocation for mobile services and the absence of end-to-end slice orchestration result in NS being unable to exert its maximum effect. Literature in [16] reveals that when NS processes a large amount of data sets, information leakage and privacy issues give rise to trust problems in information sharing between slices. The multiple machine learning methods introduced in [17] can effectively cope with dynamic resource changes, yet it is still necessary to ensure that the resource allocation among different slices does not impact QoS. Literature in [18] indicates that the RAN with performance isolation can achieve the same throughput with fewer resources, but the ability of this mechanism to counter security intrusions still requires further improvement. Literature in [19] explores the virtualization and dynamic configuration of network resources, enabling the network to possess flexibility and low-latency capabilities. In order to further improve the overall performance of network slicing, future research should focus on optimizing these aspects, especially in large-scale network deployment and cross-slicing operation scenarios, to find more efficient and secure solutions. Table II summarizes the comparison of seven articles in these aspects:

TABLE I. SUMMARY OF THE NS LITERATURE SURVEY

Reference	Summary of the Survey		
	Object	***Challenges***	***Future***
[13]	Concepts Implementation Lifecycle stages Application	Implementation complexity, interoperability, security concerns	Security exploration Efficiency Application-specific optimizations
[14]	Advantages Problems DRA Automation Configurations Security Privacy protection	Scalability Efficient resource allocation Security Privacy	Resource management with AI/ML Improved security mechanisms
[15]	Resource management Security mechanism	Ensure complete isolation Handle mobility scenarios	Simulate different mobility scenarios enhanc slice independence
[16]	Security Privacy Attack scenarios	Information leakage, securing inter-slice communication	Develop advanced AI/ML security solutions Enhance privacy protection mechanisms security solutions, enhancing privacy protection mechanisms.
[17]	Wireless access network slice resource management Dynamic resource management Isolation security	Predicting user mobility, maintaining QoS	Improving resource prediction accuracy, optimizing power allocation, enhancing mobility management
[18]	Security Attack scenarios Resource isolation	Ensuring end-to-end isolation, handling quantum security threats	Develop quantum-secure cryptographic schemes Enhance blockchain-based security measures
[19]	Cloud-native NS implementation Design principles Edge deployments	Low-latency and reliable communication, dynamic service chaining, federated resource management	Enhance slice federation Optimize service chaining Promote edge intelligent computing Integrating AI and blockchain

TABLE II. RECENT STUDIES COMPARED WITH DIFFERENT PROPERTIES

Reference	Performance Indicator			
	Effectiveness	*Efficiency*	*Scalability*	*Applicability*
[13]	High	Medium	Low	Medium
[14]	High	High	High	High
[15]	Medium	High	Medium	High
[16]	High	Medium	Medium	Low
[17]	High	High	Medium	Medium
[18]	Medium	Medium	High	Medium
[19]	High	High	High	High

NS technology has greatly optimized network resources through ML, but the current dynamic resource allocation and the isolation mechanism of dedicated slices still cannot ensure the integrity of the isolation function between NS. For example, improving interchip communication without compromising security as mentioned in [14] and [15] still lacks a comprehensive solution. In addition, as discussed in [16] and [18], applications of AI and ML in enhancing NS security have shown promise. However, the actual implementation of these solutions, especially in terms of real-time data analysis and intrusion detection, is still in the testing phase. As highlighted in [18], exploring advanced encryption methods is a promising way to address NS security vulnerabilities. In addition, although the research and development of cloud-native architecture and MEC framework for optimizing NS has begun [19], its integration with the current infrastructure is also a huge challenge to be faced.

III. NETWORK SLICING

As a novel architecture, NS can divide the physical network into multiple virtual networks. Each virtual network provides specialized services for different applications and optimizes the services continuously. NS is divided into two components, network function and Virtualized Network Function (VNF), which provide network functions for specific applications and use cases [20]. NS is supported by three main software-based network functions, namely SDN, NFV, and cloud computing[21]. When the needs of users change, cloud computing and virtualization technology can efficiently allocate shared resources to each slice of the logical network [22]. NS creates end-to-end network slices on RAN and core network to run multiple logical networks, enabling different vertical industries to run simultaneously on the same physical network infrastructure [23]. NS with complete functions can route and control packets throughout the network without being affected by other slices [24]. NS has pioneered original business conception such as Network Slicing as a Service (NSaaS) to increase revenue for mobile network operators (MNO) and hand control of networks to third parties [25].

A. Architecture

Next Generation Mobile Network (NGMN) alliance divides NS into three layers: infrastructure layer, network slice instance layer and service instance layer [26]. Fig. 1 shows the NS architecture.

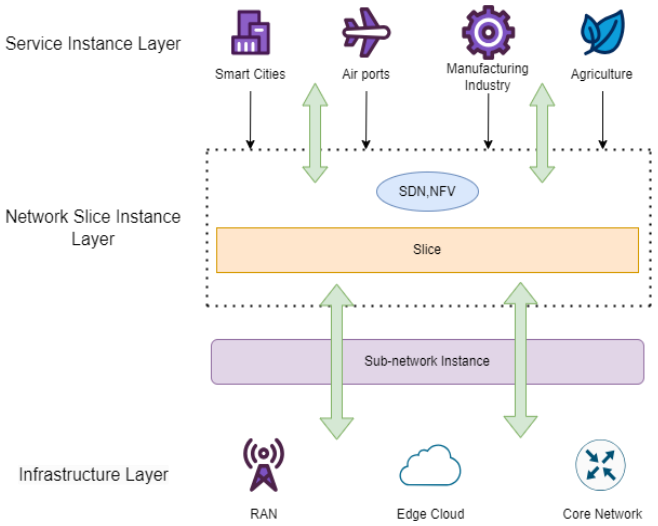


Fig. 1. NS architecture by NGMN.

1) Service instance layer: includes end-user and commercial service provided by network operators or third parties through service instances, such as applications such as virtual reality (VR), remote office, smart city, telemedicine, etc. This layer uses authentication to ensure secure access for users and devices. Distribute workloads across multiple servers to optimize resource efficiency and response time [27]. Real-time batch data processing and analysis to ensure that the data can be efficiently distributed to the terminal users. The service instance layer directly uses the network functions provided by the NS instance layer to satisfy the different needs of different applications in terms of low latency, high bandwidth, or high reliability [28]. The service instance layer includes service orchestration and management functions to ensure the running efficiency of service instances. Service orchestration manages the deployment and operation of different service instances to ensure that they cooperate. Leverage automated tools and techniques (AI/ML) to schedule resources, optimize performance, and handle failures [29].

2) Network slice instance layer: One of the core layers responsible for building virtual network instances over the top of the infrastructure layer. In this layer, each NS instance can be customized and run independently to satisfy the needs of different applications. NFV technology virtualizes the functions of traditional network devices into software modules, which optimizes the network to make it more flexible and scalable [30]. SDN and NFV are important components of the NS instance layer. SDN achieves the flexibility and dynamics of network management by separating the control plane from the data plane [31]. The control plane is managed by a centralized SDN controller. The controller communicates with network devices through an open interface to dynamically configure and manage network traffic. The data plane consists of physical and virtual network devices and performs the forwarding work of data packets. The data plane device receives the instructions issued by the control plane, and carries out the forwarding, filtering, and

processing of data packets [32]. NFV decouples the network functions and dedicated hardware devices, then runs them in the form of software on standardized hardware, thus achieving flexibility in network function deployment and management [33].

3) Infrastructure layer: Provides the physical and virtual resources required to support NS instances and service instances. This layer realizes the flexible allocation and management of resources through virtualization technology and SDN technology, which ensures the efficient operation of NS and the satisfaction of business requirements [34]. The physical hardware is the core of the infrastructure layer, with high-performance servers running VNFs and applications [35].

B. Type

NS can be classified by multiple scene: vertical and horizontal, static and dynamic, and RAN and core.

1) Classified by application scenarios: Vertical NS can establish multiple independent virtual networks can be set up on a standalone physical network infrastructure. Each slice is customized and optimized according to the different requirements of different applications. Vertical slicing allows assets to be shared between different applications and services to improve the network's service quality [36]. Each node of the network realizes the comparison capacity in a specific slice, and the traffic of different applications is isolated to complete the on-demand transmission of clients [37]. Horizontal NS distributes resources among system nodes to balance the capabilities of each node. This method provides resource isolation and edge processing [38].

2) Classified by implementation: Static NS configuration is done before the network is deployed or at an early stage of operation. Once deployed, the sliced resources will not change. The resource allocation of static slice is fixed and cannot be adjusted in real time according to the needs of applications [39]. However, its network performance and service quality are relatively stable, and its management and maintenance are also relatively simple. Therefore, it is more suitable for scenarios with stable demand such as private networks. Fixed resource allocation cannot be adjusted in real-time demand scenarios, leading to resource waste [40]. Dynamic NS can adjust resources in real time according to application requirements and network conditions, which greatly increases resource utilization and service quality [41]. It is more suitable for streaming media, healthcare, smart vehicles and other scenarios where the demand changes frequently and the flexibility and availability need to be guaranteed. However, the dynamic adjustment of resources will also bring challenges to network stability and service quality [42].

3) Classified by network architecture: RAN NS mainly manages and optimizations the resources of base station, radio spectrum, radio frequency resources and other radio access parts [43]. It can customize services for the access layer of different applications and handle mobility issues such as handover and roaming of user devices [44]. RAN NS is mostly used for some high bandwidth, low latency applications. And can connect to large-scale Internet of things like smart city sensor networks. Core NS mainly solves and optimizes the resource problems of the core network, such as data routing, session management and security control [45].

C. Lifetime

NS life cycle can be divided into four stages. The basic flow chart is shown in Fig. 2.

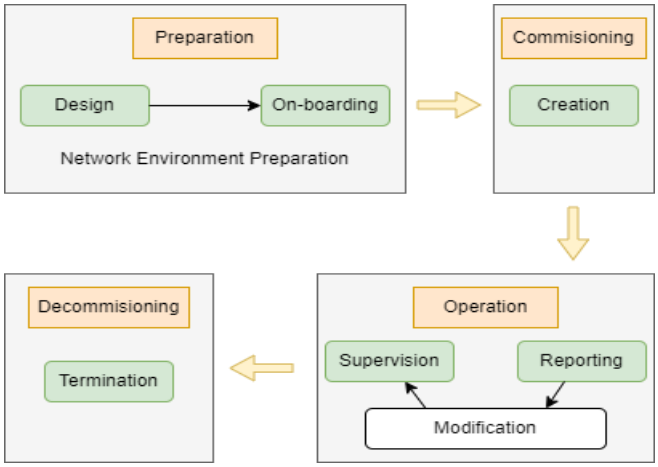


Fig. 2. NS life cycle.

1) Preparation: Network slices are planned and designed to ensure the allocation of resources and configuration. This phase needs to collect and analyze the requirements of different applications, such as bandwidth-delay reliability [46]. According to the application's requirements, the corresponding planning is made to ensure that the slice's resource allocation and service quality meet the requirements. Part of the network resources are reserved so that the dynamic allocation in the subsequent stages can proceed smoothly [47].

2) Commissioning: Setting the relevant network parameters and functions according to the specific network slice designed in the preparation phase [48]. The test verifies that the function of the slice can work properly. After ensuring that the slicing function is normal, the integration test with other services is carried out.

3) Operation: Ensure the efficient and stable operation of slices, and monitor the running status, performance indicators and service quality of slices in real-time [49]. According to the monitoring data and performance analysis results, the optimization is carried out to ensure the utilization of slice resources.

4) Decommissioning: To release the resources occupied by the network slice, clean up the data related to the deleted slice, and avoid security risks [49].

D. Resource Allocation

This section introduces two NS-based resource allocation methods.

1) An intelligent fuzzy-based system: IoT can enhance the performance of facilities which is connected by 5G. However, due to small cell sites' large-scale and unplanned deployment, switching are key issues. Due to the large number of wireless access technologies and technologies in 5G, the switching operation is complicated. In reference [50], the author proposes a fuzz-based system for 5G wireless network switching decision in view of different parameters of slicing. Two fuzzy-based switching decision models (FHDM) are constructed. In the simulation evaluation, the two models select different input parameters to see the impact on the output parameter switching decision (HD). FHDM1 selects the parameters slicing delay (SD), slicing bandwidth (SB), and slicing load (SL). FHDM2 Select Parameter slice Reliability (SR). The results show that the parameters have different influence on HD. HD values increased with the growth of SD and SL values and decreased with the increase of SB and SR values. Although FHDM2 is more complex than FHDM1, but it uses fewer input parameters, so the HD performance of FHDM2 is better. The simulation results of FHDM2 show that when the SD value is changed from 10% to 90%, and the remaining parameters are selected as 50% SR and 90% SL, the HD value increases by 50%. This suggests that when a user connects to a slice with poor service quality, switching to a better QoS is a good approach.

2) Multi-armed bandit based approach in LoRaWAN: With the increasing popularity of wireless sensor networks, LoRa WAN networks have an increasing demand for networks. Efficient resource allocation is an important difficult problem for LoRa WAN. Literature [51] introduced three schemes for how to allot the resource, and these schemes used the Multi-Armed Bandit (MAB) algorithm to maximize the packet delivery rate. UCB-MAB scheme adopts an Upper Confidence Bound (UCB) to find ways to coexist exploration and exploitation. Q-UCB-MAB scheme use Q-learning to update the Q value continuously, and it is combined with the UCB strategy to achieve deeper optimization. The third strategy of ARIMA-UCB-MAB is based on the Autoregressive Integrated Moving Average (ARIMA) model in the UCB framework to predict rewards and enhance the network's integral performance [51]. The results reveal that the three solution strategies achieve the best way to allot resource for PDR and SLA satisfaction. At the same time, with the increase of the parameters, the performance of the three schemes is optimized to a certain extent. In general, the solution strategy of ARIMA-UCB-MAB significantly outperforms the other two schemes.

IV. Technology Based on NS

Since different applications have different network requirements, NS is widely used to customize the network environment for different applications [52]. This section introduces three applications that optimize services with NS, include IoMT, IIoT and Vehicle-based system. Table III summarized technologies, existing problems and limitations of network slicing in different 5G scenarios.

A. Based on Internet of Medical Things

NS is also starting to be used in real life. With the popularity of the healthcare system, IoMT has become popular. From real-time intensive care monitoring to telemedicine, these services have highly different QoS requirements on communication networks. At present, ML/AI technology are widely used to solve these problems [53]. ML/AI analyzes data to predict patient health trends, remotely monitors patient health status, and even protects patient privacy and security. Fig. 3 shows the architecture of IoMT.

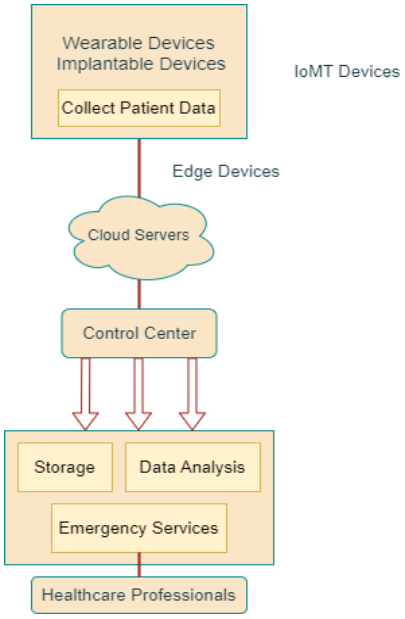


Fig. 3. IoMT architecture.

1) MT-VNE algorithm based on matching game theory: In reference [54], the authors proposed a Virtual Network Embedding solution based on two-sided matching theory (MT-VNE) to model network slices. In the NS model of medical system, each medical service is abstracted as a logical virtual network (VN), and these virtual nodes are connected by virtual links to Construct the virtual network topology. Each VN is equipped with resources that match its needs, such as processing, storage, and bandwidth resources. This innovate approach create many medical slices with different characteristics, which can be managed in one underlying network. Results show that MT-VNE performs excellently in service acceptance and resource utilization. Research will survey the impact of the underlying and virtual topology characteristics, and the future load variation of IoMT services.

2) Digital real-time health care system: During the spread of COVID-19 virus, a telemedicine system that can meet the real-time data transmission capability has become a research hotspot. But it is also a huge challenge. The authors of [55] propose an implementable architecture for digital health systems. To satisfy the needs of an end-to-end digital medical system, 5G NS is introduced to implement the system. Ensure the specificity of communication networks to satisfy the requirments for QoS in healthcare. During the architecture

design process, select an URLLC with well availability and security, guaranteed QoS, to deliver real-time patient-centered data and mobility.

The digital health systems include user plane and control plane. User plane consists of a UPF that anchors user session interconnection and traffic based on QoS which is specific. According to the data collection rules and the feasibility study of the amount of data to be deal, the dynamic slice allocation mechanism is used to accomplish this. This will ensure a commitment to patient-centered service quality. The control plane components for controlling healthcare slicing, allocating dynamic resources, and interconnecting healthcare NS with traditional telecommunications networks for voice and data communication [56]. In the future, the model will be applied to two different types of slices. Another important direction is the confidence interval of the analysis results and how to calculate the confidence interval if the analysis is integrated. The utilization of predictive advantages depends largely on the accuracy of the analysis, which also depends largely on the algorithms used to mine and process the data.

B. Based on the Industrial Internet of Things

Industrial automation is one of the most notable applications of IIoT. Many devices in the manufacturing industry require high reliability and low latency communication to ensure real-time control and monitoring of production lines [57]. Through NS, factories can create network slices dedicated to automation control, ensuring timeliness and stability of data transmission, thereby improving production efficiency and reducing downtime. Fig. 4 shows the IIoT transmission Architecture.

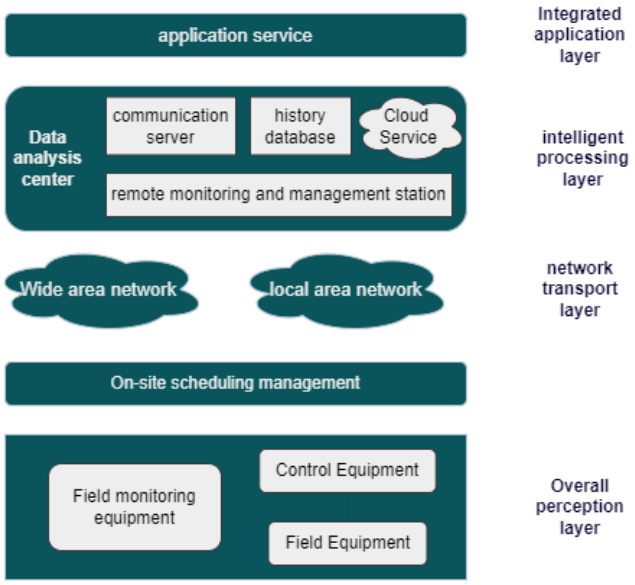


Fig. 4. IIoT architecture.

1) Resource allocation in digital siamese networks: Different services in the personalization industry place higher demands on the existing NS and resource allocation algorithms. At present, the efficiency of the algorithm is too slow and the cost is too high. In this context, the authors of [58] proposed a new network architecture in Digital Siamese Network (DTN). The architecture assists Industrial Internet of Things (IIoT) by integrating methods. The architecture is divided into three layers, three modules and two closed loops. In addition, proposes a new resource allocation system model and an solution to maximize the weighted net profit of service equilibrium. The dual-channel weighted (DCW) Critic network was introduced to realize the traffic balance and ensure the reasonable allocation of resources for personalized services. At the same time, an improved Priority Experience Replay (PER) structure was added to accelerate the convergence speed of the algorithm. This paper also proposes a PER-DCW Multi-Agent Deep Deterministic Policy Gradient (PER-DCW MADDPG) algorithm for allocating the resource in DTN-IIoT to ensure real-time operation while maintaining traffic balance.

The results show that it is necessary to ensure the service balance while dealing with the complex problem of personalized service resource allocation. The DCW network has a significant speed disadvantage, while the reliability of the network and automatic fault detection issues are also a great challenge [58]. It will continue to expand the research and use the scheme with Beyond 5G (B5G) and 6G networks. It optimizes the fault detection function of automatic network, the migration of virtual network functions and the link handover under fault conditions to make the network more reliable.

2) Non-orthogonal NS: The authors of [59] use non-orthogonal multiple access (NOMA) NS to improve the connection density of the network. In this study, the authors transformed the connectivity maximization problem into a mixed integer nonlinear programming (MINLP) problem. Subcarrier association and power allocation are jointly optimized to maximize the connectivity. The MINLP problem is first transformed into a mixed integer linear programming (MILP) problem. Then, it is further simplified into an integer linear programming (ILP) problem by formulating an effective pairing criterion. The Alternating Selection Best-Effort Pairing (AS-BEP) algorithm is used to improve the efficiency of solving the ILP problem and reduce the computational complexity. Non-orthogonal multiple access (NOMA) technology was used to share the same time-frequency resources for multiple users, and this technology improved the spectrum efficiency and network capacity.

Due to the need to support many connected devices with different service requirements, connection density and resource allocation become an important challenge. At the same time, high flexibility and high scalability are also an important requirement of industrial Internet of things for NS. In the future, simplify the complexity of MINLP algorithm to make it more suitable for real-world applications is necessary, and to continuously evaluate its feasibility and continuously optimize the algorithm in the application. Integration of NOMA and NS with other emerging technologies simultaneously, such as ML and artificial intelligence, continues to be explored to enhance decision-making processes and resource allocation strategies.

C. Vehicle-Based System

5G systems support many new things in verticals such as auto industry and power, which need more performance and cost requirements [60]. Through NS technology, vehicles can achieve real-time communication and data exchange with the cloud, to achieve intelligent navigation, remote monitoring and remote diagnosis. Fig. 5 shows the Vehicle Communication Architecture.

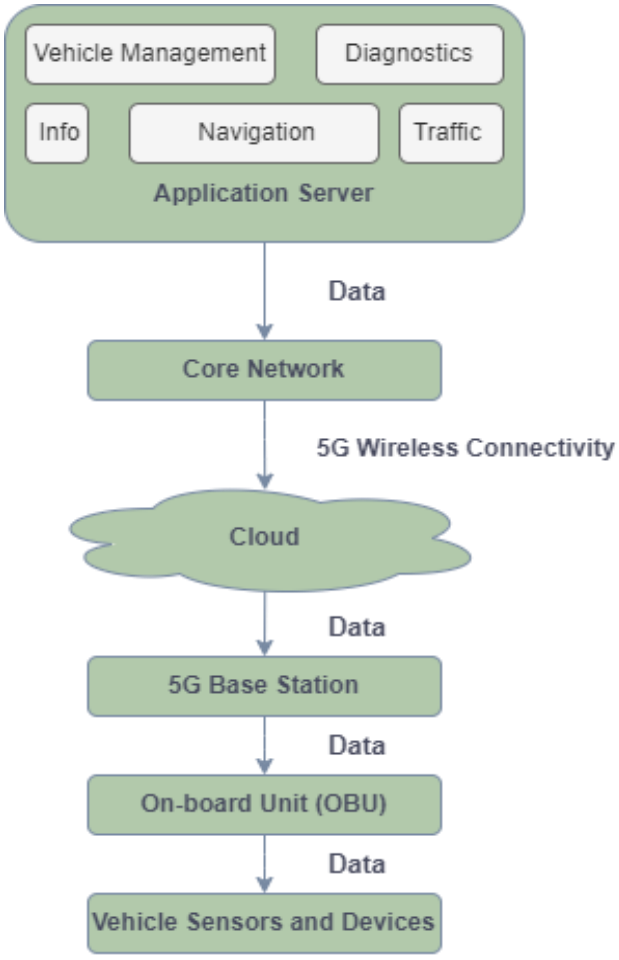


Fig. 5. Vehicle communication architecture.

1) Deep Reinforcement Learning based on Actor-Critic: A solution strategy which is appropriate for the standard 3GPP NS architecture is introduced in literature [61], which implements the Actor-Critic-based Deep Reinforcement learning (DRL) algorithm in the NS Subnet Management Function (NSSMF). The construction of this algorithm is based on the unrolling of in-vehicle applications. Because the limitation of resources, the standard slicing architecture is difficult to meet many heterogeneous service requirements without compromising the quality of service. Therefore, limited resources need to be structured in real time and efficiently to accommodate the traffic demands of different slices.

This computational approach allocates and manages resources between different generated slices in real time. In this paper, the algorithm was trained to enhance the QoS of service types in the network [61]. In the training evaluation of the computational method, the service level satisfaction of the slices is analyzed first. The results show that the satisfaction values for each slice never drop below 80% or 75%. However, the satisfaction value will decrease significantly if trained under a fixed allocation scheme. Therefore, the evaluation performance of the Actor-Critic DRL algorithm mentioned in this article is obviously better than that of the default fixed resource.

2) Vehicle-to-everything communication: A framework for V2X applications is proposed in [62]. The framework considers the coexistence of V2X communication with many other types of services to handle slice allocation. NS instances are defined within this framework. These slice instances need to interoperate and correspondence. The client instance must be configurated in each car and locked, counted, released and moved when connecting the session. The base station instance needs to accept the client request and forward it to the slice management instance; slice or slice management checks the availability of wireless resources and assigns corresponding resources to each one.

Research results show that network performance fluctuates in terms of resource consumption, switching, and blocking as the quantity of users in the region of interest changes. To provide optimal performance under a variety of network deployments and traffic conditions, resource allocation between slices must keep balanced [63].

TABLE III. NS TECHNOLOGY IN DIFFERENT APPLICATION SCENARIOS

5G Scenarios	Research Based on NS		
	Technology	***Problem***	***Limitation***
IoMT	NS ML AI	Meets different QoS require Predicts health trend of patients Monitors health status of patients remotely Protects privacy and security of patients	execution time of VNE is reduced, characteristics of underlying and virtual topology, Canges of medical service load
IIoT	NS DTN DCW PER MADDPG	Improve the reliability and low latency of real-time control and monitoring of production lines Eure service balance optisrce allocation	Resource management with AI/ML Improved security mechanisms
V2X	NS DRL MEC	Implement intelligent navigation, remote monitoring and remote diagnosis, allocates and manages resources of different slices in real time improves QoS of different service types.	Simulate different mobility scenarios enhance slice independence

V. NS Security Issues

An important part of 5G networks is to take apart network to provide different resource parts. If proper isolation is not provided for slices, the quality may be adversely impacted by flow and other factors [62]. During isolation operations, the safety of slicing becomes a central issue for most studies and studies. Softwareization using SDN and NFV in 5G networks is filling the gap in programmable control and management of network resources [64]. Table IV shows the advantages and disadvantages of each technology.

A. Software-Defined Networking

SDN technology reconstructs the traditional architecture, transforming from distributed to centralized control network architecture. SDN architecture is divided into three layers: application layer, control layer and facility layer [65]. SDN-based security threats are centered around a three-tier architecture. The first part is the data forwarding layer Switches are an important part of providing correct forwarding services for data packets at the data forward. The attacker can attach a link to one of the switch ports, threatening the end user's entry to the network [66]. The second part is the control layer translates the requirements of the SDN application layer to the SDN forwarding layer. When the controller is attacked, the whole network is affected. This is because when the switch does not receive the forwarding rule; It does not know how to forward the packet [67]. The third part is the attackers can modify the network configuration and steal network information through the application layer. This situation affects the control layer's operation and reduces the network's reliability and availability. Fig. 6 shows the SDN Security Architecture.

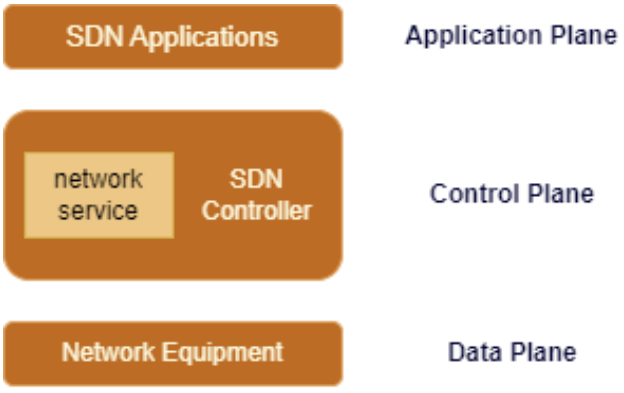


Fig. 6. SDN security architecture.

B. Network Function Virtualization

Network Function Virtualization (NFV) is a technique that virtualizes network functions using dynamic allocation to share the same facility with different functions [68]. In this technology, security problems can be caused in the following ways. When a VM allocates limited memory but unlimited resources, DoS attacks occur. Due to the complexity of integration, the integrity of the infrastructure is threatened; Resource abuse occurs when the hypervisor and scheduler fail, or when there is a lack of hypervisor isolation [69].

C. Multi-Tenant NFV/SDN

Literature [70] proposed a 5G access network security architecture supporting multi-tenant NFV/SDN. The architecture consists of security policy manager, monitoring and analysis, and Virtual Security Function (VSF). This security architecture is compatible with existing network virtualization efforts and enables fast and secure automated configuration through NFV/SDN technology. The framework mainly aims at automatic security management and VSF-based virtualization security services. The implementation of automated management relies on the service policy manager and the service monitoring and analysis policy. Virtualized security services are executed using VNS and automatically provided through the orchestrator and Virtualization Infrastructure Manager (VIM) in the NFV environment.

TABLE IV. Security Events based on different Network Slicing Technology

Technology	Research Based on NS	
	Advantage	***Limitation***
SDN	Provides centralized network control Optimized network performance Deploy applications flexibly	Controller single point of failure Complex security management Configuration challenges
NFV	Improve resource utilization Enhance infrastructure stability and resilience	High management complexity Security vulnerabilities
Hybrid NFV/SDN	Support resource sharing Improve operation and maintenance efficiency Automate security configuration	Security management and monitoring mechanism is complex Cost and resource investment are high

VI. Discussion

Through the investigation of existing research found that NS improves network efficiency and service quality by dividing logical networks, to realize the optimization of latency, resource management and security privacy. Current research trends focus on the integration of AI and ML to automate network management, resource prediction, and allocation. NS technology will enhance its application in large-scale networks by means of intelligent and automated management.

The isolation effect between slices poses a threat when resources are allocated. The defect of slice isolation will cause mutual interference between slices, which will affect the application of NS in the high movement scene of 5G. In addition, most current studies mainly rely on simulation and experimental environments, and lack the operability of resource allocation and isolation methods in complex network environments. Especially in applications such as IoMT, IIoT and in-vehicle networking, the reliability and latency of NS do not fully meet the demand. Secondly, because the challenges of 5G slicing in dynamic service chain management, edge

computing and joint resource management have not been fully solved, it is difficult to achieve efficient use of resources and continuous guarantee of service quality in practical applications.

VII. Challenges and Future Directions

NS technology faces major challenges such as efficiently and dynamically allocating resources according to demand, ensuring isolation effects, and avoiding interference. In this environment, accurate prediction of user mobility and demand to maintain quality of service and optimize resources is necessary to address these challenges [71]. Data leakage and unauthorized access in the isolation process will bring security risks to the slice information. Advanced encryption technologies such as blockchain can solve security problems such as location tracking and impersonation attacks [72]. Interoperability between different slices and integration with existing network infrastructure causes compatibility issues between vendors and carriers [73]. Developing common standards and protocols to ensure that each slice scales independently and maintains quality of service during peak load can effectively solve this problem. Orchestration and management frameworks can also be used to efficiently handle the dynamic scaling capabilities of slicing, which further optimizes low latency and reliability between services [74].

A. Challenges and Requirements

NS technology needs to allocate resources efficiently and dynamically according to demand, and ensure the isolation effect between slices [75]. This is a high requirement for predicting demand and allocating application resources. Accurately predicting user mobility and demand can also optimize the quality of service maintenance and the allocation of resources [76]. Data breaches and unauthorized access can create security issues in the isolation process.

B. Future Work

In the future, it is necessary to strengthen the application of AI/ML in resource management and develop more efficient dynamic service chain management algorithms to improve the accuracy of resource prediction [77]. Advanced security solutions are implemented using quantum secure encryption schemes and blockchain technology to enhance data protection against unauthorized access and automate SLA management [78]. Continue to explore AI/ ML-based security measures to detect and mitigate sophisticated attacks. At the same time, common NS standards and protocols are established to promote the interoperability between slicing and seamless integration with the existing network infrastructure [79]. Combine edge computing with NS to establish low-latency, efficient services, and optimize resource utilization and service delivery through AI-driven edge computing solutions [80].

VIII. Conclusion

This article explores NS technology in depth and demonstrates its strong potential for use in different 5G application scenarios. The analysis of NS architecture, resource allocation mechanism and application examples reveals the significant advantages of NS technology in the optimization of resource utilization, network latency, quality of service and security in IoMT and other fields. This research also found that the current NS technology has great challenges in cross-slice interoperability, security isolation and dynamic resource allocation. In the future, the optimization of NS architecture and the strengthening of security measures should be strengthened. Through the optimization of network slice management to improve resource utilization and network transmission performance. Use AI/ML to achieve intelligent and automated NS management, while introducing more advanced security encryption measures like blockchain and quantum encryption technology. As challenges continue to be discovered and solved, NS technology will continue to drive the digitization and intelligence of 5G applications.

Acknowledgment

This work was supported in part by the Universiti Kebangsaan Malaysia Fundamental Research Grant Scheme (FRGS) from the Ministry of Higher Education with the code: FRGS/1/2023/ICT07/UKM/02/1 and FRGS/1/2022/ICT11/UKM/02/1.

References

[1] Ssengonzi, C., Kogeda, O.P. & Olwal, T.O. 2022. A survey of deep reinforcement learning application in 5G and beyond NS and virtualization. *Array* 14: 100142.

[2] Fadlullah, Z. M., Mao, B., & Kato, N. (2022). Balancing QOS and security in the edge: existing practices, challenges, and 6G opportunities with ML. *IEEE Communications Surveys and Tutorials/IEEE Communications Surveys and Tutorials*, *24*(4), 2419–2448. https://doi.org/10.1109/comst.2022.3191697

[3] Ghadi, Y.Y., Shah, S.F.A., Mazhar, T., Shahzad, T., Ouahada, K. & Hamam, H. 2024. Enhancing patient healthcare with mobile edge computing and 5G: challenges and solutions for secure online health tools. *Journal of Cloud Computing* 13(1).

[4] Escolar, A.M., Wang, Q. & Calero, J.M.A. 2024. Enhancing honeynet-based protection with NS for massive Pre-6G IoT Smart Cities deployments. *Journal of Network and Computer Applications*: 103918.

[5] Upadhyay, D., Gupta, A., Mohd, N. & Pant, B. 2023. A review of NS based 5G. *AIP Conference Proceedings*.

[6] Siddiqui, M. U. A., Qamar, F., Ahmed, F., Nguyen, Q. N., & Hassan, R. (2021). Interference Management in 5G and Beyond Network: Requirements, challenges and future directions. *IEEE Access*, *9*, 68932–68965.

[7] Subedi, P., Alsadoon, A., Prasad, P. W. C., Rehman, S. U., Giweli, N., Imran, M., & Arif, S. (2021). NS: a next generation 5G perspective. *EURASIP Journal on Wireless Communications and Networking*, *2021*(1). https://doi.org/10.1186/s13638-021-01983-7

[8] Hassan, S. M. H. U., Brennan, A., Muntean, G. M., & McManis, J. (2024, June). NSM 2: Network Slice Management and Monitoring Using Machine Learning For AR/VR Applications. In *2024 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)* (pp. 1-7). IEEE.

[9] Ramesh, B., Changala, R., Kalangi, P. K., & Bhavsingh, M. (2024). Optimizing 6G Network Slicing with the EvoNetSlice Model for Dynamic Resource Allocation and Real-Time QoS Management. *International Research Journal of Multidisciplinary Technovation*, *6*(3), 325-340.

[10] Georgieva, P., Vlahov, A., Poulkov, V., & Manolova, A. (2024, July). A Machine Learning Approach for Network Slice Selection. In *2024 59th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST)* (pp. 1-5). IEEE.

[11] Blanco, L., Kukliński, S., Zeydan, E., Rezazadeh, F., Chawla, A., Zanzi, L., ... & Mangues, J. (2023). AI-driven framework for scalable management of network slices. *IEEE Communications Magazine*, *61*(11), 216-222.

[12] Saravanan, N., & Mathina, P. A. (2024, August). SDN-Enabled Networks: A Deeper Learning-Based Echo State Neural Network for Cyber Threat Detection. In *2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT)* (Vol. 1, pp. 1356-1360). IEEE.

[13] Wijethilaka, S. & Liyanage, M. 2021a. Survey on NS for Internet of Things Realization in 5G networks. *IEEE Communications Surveys and Tutorials/IEEE Communications Surveys and Tutorials* 23(2): 957–994.

[14] Wijethilaka, S. & Liyanage, M. 2021b. Realizing Internet of Things with NS: Opportunities and Challenges. *IEEE*.

[15] Alotaibi, D. 2021. Survey on Network slice isolation in 5G Networks: Fundamental Challenges. *Procedia Computer Science* 182: 38–45.

[16] De Alwis, C., Porambage, P., Dev, K., Gadekallu, T.R. & Liyanage, M. 2024. A survey on NS Security: attacks, challenges, solutions and research directions. *IEEE Communications Surveys and Tutorials/IEEE Communications Surveys and Tutorials*: 1.

[17] Azimi, Y., Yousefi, S., Kalbkhani, H. & Kunz, T. 2022. Applications of ML in Resource Management for RAN-Slicing in 5G and Beyond Networks: A survey. *IEEE Access* 10: 106581–106612.

[18] Salahdine, F., Liu, Q. & Han, T. 2022. Towards secure and intelligent NS for 5G networks. IEEE Open Journal of the Computer Society 3: 23–38.

[19] Shah, S.D.A., Gregory, M.A. & Li, S. 2021. Cloud-Native NS using software defined Networking based Multi-Access Edge Computing: A survey. IEEE Access 9: 10903–10924.

[20] Zhang, S. 2019. An overview of NS for 5G. IEEE Wireless Communications 26(3): 111–117.

[21] Shinde, S. S., Marabissi, D., & Tarchi, D. (2021). A network operator-biased approach for multi-service network function placement in a 5G NS architecture. *Computer Networks*, *201*, 108598. https://doi.org/10.1016/j.comnet.2021.108598

[22] Sweidan, Z., Islambouli, R. & Sharafeddine, S. 2020. Optimized flow assignment for applications with strict reliability and latency constraints using path diversity. Journal of Computational Science 44: 101163.

[23] Motalleb, M. K., Shah-Mansouri, V., Parsaeefard, S., & López, O. L. A. (2023). Resource allocation in an open RAN system using NS. *IEEE Transactions on Network and Service Management/IEEE eTransactions on Network and Service Management*, *20*(1), 471–485. https://doi.org/10.1109/tnsm.2022.3205415

[24] Escolar, A. M., Alcaraz-Calero, J. M., Salva-Garcia, P., Bernabe, J. B., & Wang, Q. (2021). Adaptive NS in Multi-Tenant 5G IoT networks. IEEE Access, 9, 14048–14069. https://doi.org/10.1109/access.2021.3051940

[25] Cardenas, A., Fernandez, D., Lentisco, C. M., Moyano, R. F., & Bellido, L. (2021). Enhancing a 5G NS management model to improve the support of mobile virtual network operators. *IEEE Access*, *9*, 131382–131399. https://doi.org/10.1109/access.2021.3114645

[26] Mohammed, C.M. & Shaikhah, S.K. 2022. A Survey and Analysis of Architecture and Models of NS in 5G. IEEE.

[27] Li, X., Ni, R., Chen, J., Lyu, Y., Rong, Z., & Du, R. (2020). End-to-End NS in radio access network, transport network and core network domains. *IEEE Access*, *8*, 29525–29537. https://doi.org/10.1109/access.2020.2972105

[28] Rahmanian, G., Shahhoseini, H. S., & Pozveh, A. H. J. (2021). A Review of NS in 5G and Beyond: Intelligent Approaches and Challenges. IEEE. https://doi.org/10.23919/ituk53220.2021.9662097

[29] Martins, J. S. B., Carvalho, T. C., Moreira, R., Both, C., Donatti, A., Corrêa, J. H., Suruagy, J. A., Corrêa, S. L., Abelem, A. J. G., Ribeiro, M. R. N., Nogueira, J. M. S., Magalhães, L. C. S., Wickboldt, J., Ferreto, T. C., Mello, R., Pasquini, R., Schwarz, M., Sampaio, L. N., Macedo, D. F., . . . Silva, F. O. (2023). Enhancing NS architectures with ML, security, sustainability and experimental networks integration. IEEE Access, 11, 69144–69163. https://doi.org/10.1109/access.2023.3292788

[30] Borsatti, D., Grasselli, C., Contoli, C., Micciullo, L., Spinacci, L., Settembre, M., Cerroni, W., & Callegati, F. (2023). Mission critical communications support with 5G and NS. *IEEE Transactions on Network and Service Management/IEEE eTransactions on Network and Service Management*, *20*(1), 595–607. https://doi.org/10.1109/tnsm.2022.3208657

[31] Park, K., Sung, S., Kim, H. & Jung, J.-I. 2023. Technology trends and challenges in SDN and service assurance for end-to-end NS. *Computer Networks* 234: 109908.

[32] Alotaibi, D., Thayananthan, V., & Yazdani, J. (2021). The 5G NS using SDN based technology for managing network traffic. *Procedia* Computer Science, 194, 114–121. https://doi.org/10.1016/j.procs.2021.10.064

[33] Liao, C., Lin, F. J., & Sato, Y. (2020). Evaluating NFV-enabled NS for 5G Core. *IEEE*. https://doi.org/10.23919/apnoms50412.2020.9236978

[34] Debbabi, F., Jmal, R., Fourati, L. C., & Ksentini, A. (2020b). Algorithmics and Modeling Aspects of NS in 5G and Beyonds Network: survey. *IEEE Access*, *8*, 162748–162762. https://doi.org/10.1109/access.2020.3022162

[35] Kourtis, M., Anagnostopoulos, T., Kuklilski, S., Wierzbicki, M., Oikonomakis, A., Xilouris, G., Chochliouros, I. P., Yi, N., Kostopoulos, A., Tomaszewski, L., Sarlas, T., & Koumaras, H. (2020). 5G NS Enabling Edge Services. *IEEE*. https://doi.org/10.1109/nfv-sdn50289.2020.9289880

[36] Xie, M., Gonzalez, A. J., Gronsund, P., Lonsethagen, H., Waldemar, P., Tranoris, C., Denazis, S., & Elmokashfi, A. (2022). Practically Deploying Multiple Vertical Services into 5G Networks with NS. *IEEE Network*, *36*(1), 32–39. https://doi.org/10.1109/mnet.001.2100361

[37] Zahoor, S., Ahmad, I., Othman, M.T.B., Mamoon, A., Rehman, A.U., Shafiq, M. & Hamam, H. 2022. Comprehensive analysis of NS for the developing commercial needs and networking challenges. Sensors 22(17): 6623.

[38] Hegde, P., & Meena, S. (2021). A survey on 5G NS-Epitome and opportunities for a novice. *IEEE*. https://doi.org/10.1109/icccnt51525.2021.9579745

[39] Goncalves, D., Puliafito, C., Mingozzi, E., Rana, O., Bittencourt, L., & Madeira, E. (2020). Dynamic NS in Fog Computing for Mobile Users in MobFogSim. *IEEE*. https://doi.org/10.1109/ucc48980.2020.00042

[40] Reyhanian, N., & Luo, Z. (2022). Data-Driven adaptive NS for Multi-Tenant networks. *IEEE Journal of Selected Topics in Signal Processing*, *16*(1), 113–128. https://doi.org/10.1109/jstsp.2021.3127796

[41] Chochliouros, I. P., Spiliopoulou, A. S., Lazaridis, P. I., Dardamanis, A., Zaharis, Z. D., & Kostopoulos, A. (2020). Dynamic NS: challenges and opportunities. In *IFIP advances in information and communication technology* (pp. 47–60). https://doi.org/10.1007/978-3-030-49190-1_5

[42] Feng, J., Pei, Q., Yu, F. R., Chu, X., Du, J., & Zhu, L. (2020). Dynamic NS and resource allocation in mobile edge computing systems. *IEEE Transactions on Vehicular Technology*, *69*(7), 7863–7878. https://doi.org/10.1109/tvt.2020.2992607

[43] Kourtis, M., Anagnostopoulos, T., Kuklilski, S., Wierzbicki, M., Oikonomakis, A., Xilouris, G., Chochliouros, I. P., Yi, N., Kostopoulos, A., Tomaszewski, L., Sarlas, T., & Koumaras, H. (2020b). 5G NS Enabling Edge Services. *IEEE*. https://doi.org/10.1109/nfv-sdn50289.2020.9289880

[44] Hamdi, W., Ksouri, C., Bulut, H. & Mosbah, M. 2024. NS Based Learning Techniques for IoV in 5G and Beyond Networks. *IEEE Communications Surveys and Tutorials/IEEE Communications Surveys and Tutorials*: 1.

[45] Arteaga, C. H. T., Ordonez, A., & Rendon, O. M. C. (2020). Scalability and performance analysis in 5G core NS. *IEEE Access*, *8*, 142086–142100. https://doi.org/10.1109/access.2020.3013597

[46] Drif, Y., Chaput, E., Lavinal, E., Berthou, P., Jou, B. T., Gremillet, O., & Arnal, F. (2020). An extensible NS framework for satellite integration into 5G. *International Journal of Satellite Communications and Networking*, *39*(4), 339–357. https://doi.org/10.1002/sat.1387

[47] Alex, G., Francesco, T., Stuart, C., Christian, R., & Joan, S. (2020). Slicing 5G networks: an architectural survey. *Wiley 5G Ref*, 1–41. https://doi.org/10.1002/9781119471509.w5gref095

[48] Vittal, S., Sarkar, S., & Franklin, A. A. (2023). Revamping the resilience and high availability of 5G core for 6G ready network slices. *IEEE Transactions on Network and Service Management/IEEE eTransactions on Network and Service Management*, 1. https://doi.org/10.1109/tnsm.2023.3348137

[49] Montero, R., Agraz, F., Pagès, A., & Spadaro, S. (2020). Enabling Multi-segment 5G Service Provisioning and Maintenance through NS.

Journal of Network and Systems Management, *28*(2), 340–366. https://doi.org/10.1007/s10922-019-09509-9

[50] Ampririt, P., Higashi, S., Qafzezi, E., Ikeda, M., Matsuo, K. & Barolli, L. 2023. An intelligent fuzzy-based system for handover decision in 5G-IoT networks considering NS and SDN technologies. Internet of Things 23: 100870.

[51] Mardi, F.Z., Hadjadj-Aoul, Y., Bagaa, M. & Benamar, N. 2024. Resource allocation for LoRaWAN NS: Multi-Armed Bandit-based approaches. Internet of Things: 101195.

[52] Gur, G., Kalla, A., De Alwis, C., Pham, Q., Ngo, K., Liyanage, M., & Porambage, P. (2022). Integration of ICN and MEC in 5G and Beyond networks: mutual benefits, use cases, challenges, standardization, and future research. *IEEE Open Journal of the Communications Society*, *3*, 1382–1412. https://doi.org/10.1109/ojcoms.2022.3195125

[53] Dehkordi, I.F., Manochehri, K. & Aghazarian, V. 2023. Internet of Things (IoT) Intrusion Detection by ML (ML): a review. *Asia-Pacific Journal of Information Technology and Multimedia* 12(01): 13–38.

[54] Tian, C., Cao, H., Garg, S., Kaddoum, G., Hassan, M.M. & Xie, J. 2023. 5G in healthcare: Matching game-empowered intelligent medical NS. Alexandria Engineering Journal /Alexandria Engineering Journal 77: 95–107.

[55] Jain, H., Chamola, V., Jain, Y. & Naren, N. 2021. 5G network slice for digital real-time healthcare system powered by network data analytics. Internet of Things and Cyber-physical Systems 1: 14–21.

[56] Aman, A.H.M., Hassan, W.H., Sameen, S., Attarbashi, Z.S., Alizadeh, M. & Latiff, L.A. 2021b. IoMT amid COVID-19 pandemic: Application, architecture, technology, and security. *Journal of Network and Computer Applications* 174: 102886.

[57] Hassan, R., Qamar, F., Hasan, M.K., Aman, A.H.M. & Ahmed, A.S. 2020. Internet of Things and its Applications: A Comprehensive survey. *Symmetry* 12(10): 1674.

[58] Tang, L., Du, Y., Liu, Q., Li, J., Li, S. & Chen, Q. 2023. Digital-Twin-Assisted resource allocation for NS in industry 4.0 and beyond using distributed deep reinforcement learning. *IEEE Internet of Things Journal* 10(19): 16989–17006.

[59] Yin, B., Tang, J. & Wen, M. 2023. Connectivity maximization in Non-Orthogonal NS enabled industrial Internet-of-Things with multiple services. *IEEE Transactions on Wireless Communications* 22(8): 5642–5656.

[60] Sohaib, R.M., Onireti, O., Sambo, Y. & Imran, M.A. 2021. NS for beyond 5G systems: An overview of the smart port use case. Electronics 10(9): 1090.

[61] Tairq, M.A., Saad, M.M., Khan, M.T.R., Seo, J. & Kim, D. 2023. DRL-based resource management in NS for vehicular applications. ICT Express 9(6): 1116–1121.

[62] Zamfirescu, C., Iugulescu, R., Crăciunescu, R., Vulpe, A., Li, F.Y. & Halunga, S. 2024. Network slice allocation for 5G V2X networks: A case study from framework to implementation and performance assessment. Vehicular Communications 45: 100691.

[63] Barakabitze, A.A., Ahmad, A., Mijumbi, R. & Hines, A. 2020. 5G NS using SDN and NFV: A survey of taxonomy, architectures and future challenges. Computer Networks 167: 106984.

[64] Abood, M.J.K. & Abdul-Majeed, G.H. 2023. Classification of NS threats based on slicing enablers: A survey. International Journal of Intelligent Networks 4: 103–112.

[65] Kazmi, S.H.A., Qamar, F., Hassan, R., Nisar, K. & Chowdhry, B.S. 2023. Survey on Joint Paradigm of 5G and SDN Emerging Mobile Technologies: architecture, security, challenges and research Directions. *Wireless Personal Communications* 130(4): 2753–2800.

[66] Barmpounakis, S., Maroulis, N., Papadakis, M., Tsiatsios, G., Soukaras, D., & Alonistioti, N. (2020). NS - enabled RAN management for 5G: Cross layer control based on SDN and SDR. *Computer Networks*, *166*, 106987. https://doi.org/10.1016/j.comnet.2019.106987

[67] Nisar, K., Jimson, E.R., Hijazi, M.H.A., Welch, I., Hassan, R., Aman, A.H.M., Sodhro, A.H., Pirbhulal, S. & Khan, S. 2020. A survey on the architecture, application, and security of software defined networking: Challenges and open issues. *Internet of Things* 12: 100289.

[68] Alkhafaji, A. R., & Al-Turaihi, F. S. (2021). Multi-Layer NS and Resource Allocation Scheme for Traffic-aware QoS ensured SDN/NFV-5G Network. *IEEE*. https://doi.org/10.1109/bicits51482.2021.9509901

[69] Shu, Z., & Taleb, T. (2020). A novel QOS framework for NS in 5G and beyond networks based on SDN and NFV. *IEEE Network*, *34*(3), 256–263. https://doi.org/10.1109/mnet.001.1900423

[70] Siddiqui, Escalona, E., Trouva, E., Kourtis, M.A., Kritharidis, D., Katsaros, K., Spirou, S., Canales, C. & Lorenzo, M. 2016. Policy based virtualised security architecture for SDN/NFV enabled 5G access networks. IEEE.

[71] Vittal, S., & A, A. F. (2021). Self Optimizing NS in 5G for Slice Isolation and High Availability. *IEEE*. https://doi.org/10.23919/cnsm52442.2021.9615546

[72] Abdulqadder, I. H., & Zhou, S. (2022). SliceBlock: Context-Aware authentication handover and secure NS using DAG-Blockchain in Edge-Assisted SDN/NFV-6G environment. *IEEE Internet of Things Journal*, *9*(18), 18079–18097. https://doi.org/10.1109/jiot.2022.3161838

[73] Zhong, J., Chen, C., Qian, Y., Bian, Y., Huang, Y., & Bie, Z. (2024). Secure and scalable NS with Plug-and-Play support for power distribution system communication networks. *IEEE Internet of Things Journal*, *11*(12), 22036–22053. https://doi.org/10.1109/jiot.2024.3378770

[74] Ji, L., He, S., Wu, W., Gu, C., Bi, J., & Shi, Z. (2022). Dynamic NS orchestration for remote adaptation and configuration in industrial IoT. *IEEE Transactions on Industrial Informatics*, *18*(6), 4297–4307. https://doi.org/10.1109/tii.2021.3131355

[75] Zhou, J., Zhao, W., & Chen, S. (2020). Dynamic network slice scaling assisted by prediction in 5G network. *IEEE Access*, *8*, 133700–133712. https://doi.org/10.1109/access.2020.3010623

[76] Abood, M. S., Wang, H., He, D., Fathy, M., Rashid, S. A., Alibakhshikenari, M., Virdee, B. S., Khan, S., Pau, G., Dayoub, I., Livreri, P., & Elwi, T. A. (2023). An LSTM-Based NS Classification Future Predictive framework for optimized resource allocation in C-V2X. *IEEE Access*, *11*, 129300–129310. https://doi.org/10.1109/access.2023.3332225

[77] Yang, L., Jia, J., Lin, H., & Cao, J. (2023). Reliable dynamic service chain scheduling in 5G networks. *IEEE Transactions on Mobile Computing*, *22*(8), 4898–4911. https://doi.org/10.1109/tmc.2022.3157312

[78] He, Y., Zhang, C., Wu, B., Yang, Y., Xiao, K., & Li, H. (2023). Cross-Chain Trusted Service Quality Computing Scheme for Multichain-Model-Based 5G NS SLA. *IEEE Internet of Things Journal*, *10*(14), 12126–12139. https://doi.org/10.1109/jiot.2021.3132388

[79] Yan, Y., Beldachi, A. F., Nejabati, R., & Simeonidou, D. (2020). P4-enabled Smart NIC: enabling sliceable and Service-Driven optical data centres. *Journal of Lightwave Technology*, *38*(9), 2688–2694. https://doi.org/10.1109/jlt.2020.2966517

[80] Feng, J., Pei, Q., Yu, F. R., Chu, X., Du, J., & Zhu, L. (2020b). Dynamic NS and resource allocation in mobile edge computing systems. *IEEE Transactions on Vehicular Technology*, *69*(7), 7863–7878. https://doi.org/10.1109/tvt.2020.2992607

Real Time Object Detection for Sustainable Air Conditioner Energy Management System

Chang Shi Ying[1], C. PuiLin[2]
Student-Faculty of Data Science and Information Technology, INTI International University, Malaysia[1]
Senior Lecturer-Faculty of Data Science and Information Technology, INTI International University, Malaysia[2]

***Abstract*—Air conditioning has become indispensable for maintaining human comfort, especially during hot weather, as people rely on it to stay cool indoors. However, the long-term and uncontrolled use of air conditioners has significantly contributed to climate change and environmental degradation. The extensive use of air conditioners releases more carbon dioxide, a greenhouse gas, into the atmosphere, exacerbating global warming and leading to adverse climate impacts. The proposed sustainable air conditioning energy management system aims to address this issue by optimising air conditioner use while minimising its environmental footprint and mitigating climate change. Current air conditioning systems in offices, buildings, and homes typically rely on fixed temperature settings, leading to excessive energy consumption and increased greenhouse gas emissions. Existing solutions, such as fixed timers, manual timer settings, and physical controllers, are ineffective as they cannot dynamically respond to changes in environmental conditions, such as room occupancy and activity levels, resulting in significant inefficiencies and environmental hazards. To overcome these limitations, the proposed system introduces an innovative solution using software engineering technology, specifically real-time object detection, to control air conditioning energy usage. This approach redefines air conditioning management by allowing the system to dynamically adapt to room occupancy, environmental factors, and activity levels, ensuring the right amount of cooling is delivered at the right time. This method represents a concrete and effective response to climate change challenges and demonstrates a commitment to creating a sustainable and environmentally responsible future.**

***Keywords*—Deep learning; energy consumption; energy efficiency; global warming; climate change; real-time object detection; Air conditioner optimization; smart meter; environmental footprint; climate action**

I. Introduction

Climate change affects every country globally, disrupting economies and impacting vulnerable populations. Climate change affects weather patterns, rising sea levels and more frequent extreme weather events. Notably, a critical concern within this context is the high energy consumption levels and carbon emissions. Sustainability has emerged as a paramount human priority in the increasingly advanced world. Taking this into account, the United Nations adopted the Sustainable Development Goals (SDGs) in 2015 to protect the planet [1].

Sustainable development and reducing energy consumption are two of the project's main objectives. Therefore, the project aligns with SDG 13, which addresses climate change, and SDG 11, which aims to achieve sustainable cities and communities. Goal 13 emphasizes the urgency of acting against climate change, while Goal 11 aims to create inclusive, safe, resilient, and sustainable cities [2]. To achieve these goals, the project sets out to develop a sustainable air conditioning system that incorporates cutting-edge software engineering technology, specifically real-time object detection.

Driven by artificial intelligence (AI) and machine learning, real-time object detection has made significant progress, especially in the field of deep learning. This advanced technology enables computers to identify and locate objects within images or live video feeds swiftly and accurately. Deep learning is a subset of machine learning using convolutional neural networks inspired by the architecture of the human brain at its core. By training these deep neural networks on extensive datasets, computers gain the ability to discern intricate patterns and features, enabling them to recognize and categorize objects in real-time scenarios.

By incorporating real-time object detection, the air conditioner energy management system adapts dynamically to the environment, optimizing energy usage with precision. Unlike traditional technology, the system can automatically adjust the temperature based on the activity levels and the number of occupants in a room. This will ensure that air conditioning energy is not wasted and protect the environment by optimizing energy use. With real-time and remote-control capabilities, the system provides users with more convenience and control. Users can easily adjust temperature settings and ensure optimal energy usage, especially when the room is unoccupied. This user-friendly solution encourages sustainable choices effortlessly.

This project aims to create innovative environmental sustainability solutions and embody buildings by solving the issue of climate change. The objective of this project is to improve existing solutions that have limited function. By optimizing energy usage, detecting human activity and occupancy, and utilizing hardware like Cameras, Smart Sensors, and Smart Remote Controllers, this project aims to use technology to build a more sustainable future.

A. Problem Statement

The current system has several problems that must be rectified. Since it controls the temperature of the air conditioner largely with an infrared remote control, its portability and accessibility may be limited as a result. Furthermore, the focus of this research is on improving user comfort rather than improving energy efficiency, which is a vital component of any smart home system. In the second article, the system's security layer is entirely made up of authentication and authorization

functions. Furthermore, it is only compatible with the Android operating system, which implies that only Android users can use it. Third, the existing air conditioner energy management system is based on an outdated algorithm known as YOLOv3 (You Only Look Once), which lacks the enhancements found in the most recent version.

Furthermore, the existing system's technique for controlling air conditioners is limited to counting the number of persons in a room and is unable to recognize individual room activities for the purpose of providing precise temperature adjustment. There is also a method known as deep learning detection, which is based on the behavior of a single resident. It is vital to consider the presence of multiple occupants to increase overall accuracy and ability. These constraints highlight the need for an improved air conditioning management system capable of overcoming these constraints while still ensuring portability, security, accessibility, and efficacy. The proposed improvements are intended to reduce air conditioner energy use for a wide variety of users, including those in homes, businesses, and buildings.

B. Research Objectives

This research paper seeks to fulfil its objectives by addressing three key questions. Firstly, it aims to study existing air conditioning systems to understand how real-time object detection and energy management can be applied within these systems. This involves a thorough analysis of current technologies and methodologies to identify opportunities for improvement. Secondly, the research focuses on designing and developing a sustainable air conditioner energy management system that not only reduces energy consumption but also promotes a more sustainable environment. Lastly, the paper will conduct rigorous testing to ensure the effectiveness of the newly developed air conditioner energy management system in real-world scenarios. This testing is crucial to validate the system's ability to optimize energy consumption, thereby contributing to a reduction in greenhouse gas emissions and fostering a more sustainable future.

II. Literature Review

The Air Conditioner (AC) system is essential for providing human comfort, yet it also contributes to the rising outdoor temperatures seen because of climate change. This creates a concerning and self-perpetuating cycle, with the CO2 emissions from AC units only exacerbating the issue. Considering this, numerous researchers have proposed solutions to reduce the contribution of AC systems to climate change. Thus, this literature review explores the approaches taken in prior research on Air Conditioner Energy Management Systems.

The current air conditioner energy management system has integrated an IoT Smart House System emphasizing the efficient management of indoor environmental conditions to enhance occupants' thermal comfort during leisure activities. It involves analyzing indoor environmental data by sensors for CO2 and particulate matter (PM) and carefully selecting measurement nodes to ensure system stability and reliability. It uses an infrared remote control to control the temperature of the air conditioner. The focus of this system is the application of IoT and user comfort [3].

Another current system, HEMS-IoT (Home Energy Management Systems-IoT), integrates big data and machine learning to manage home energy for comfort, safety, and savings. Machine learning algorithms, such as J48, are used to analyze energy consumption and user behavior patterns, while RuleML and Apache Mahout help create energy-saving recommendations based on user preferences. A case study validates HEMS-IoT's effectiveness in reducing energy consumption while maintaining home comfort. The security layer of the HEMS-IoT ensures the confidentiality and integrity of data by employing two fundamental security components, which are authentication and authorization. While at the presentation layer, HEMS-IoT is working on the Android operating system, which is a constraint to other users such as Apple iOS [4].

Another current solution for reducing energy consumption and sustainability in smart buildings is integrating deep learning and the Internet of Things (IoT). Recognizing the significance of HVAC systems in energy consumption, Elsisi et al. proposed a deep learning-based people detection system employing YOLOv3 to count occupants within specific areas of smart buildings accurately. This innovative system optimizes the operation of air conditioners based on real-time occupancy data, contributing to enhanced energy efficiency. Moreover, the IoT platform's dashboard facilitates data visualization and informed decision-making regarding energy consumption [5]. The current air conditioner energy management system relies on an outdated YOLO deep learning algorithm (YOLOv3) for object detection. However, YOLOv8, the latest version, offers enhancements such as anchor-free, improved CNN architecture, more efficient, and improved average precision score [6].

Furthermore, an existing system employs a data-driven deep learning method employing a convolutional neural network (CNN) model to detect various occupancy activities. The performance of this deep learning framework was assessed in an office space within a case study building, initially identifying activities like 'standing,' 'sitting,' 'walking,' 'napping,' and 'none,' even for detecting no occupancy. This approach achieved an average detection accuracy of 80.62% during live testing, highlighting its potential for real-time monitoring. The Deep Learning Influenced Profile (DLIP) was also evaluated in estimating sensible heat gains driven by occupancy and CO2 concentrations [7].

Machine and deep learning methods are highly effective for building energy prediction due to their advantages over traditional models, including reduced data requirements and faster model development. Machine and deep learning excel in energy forecasting, thermal comfort prediction, and occupancy detection. Deep understanding is gaining popularity in these fields thanks to increased data availability and improved algorithms. Some studies suggest combining deep and traditional machine learning techniques can further enhance application performance [8].

III. Methodology

This real-time object detection of a sustainable air conditioning energy management system will use a life cycle (SDLC) as a prototype of the research method. With this approach, the research process is considered through various

stages, each of which contributes to the development of this system. Using the SDLC framework can make research objectives more precise and more organized.

Fig. 1 shows the phase for this research. The research methodology includes requirement gathering, design system architecture, building prototype, and testing.

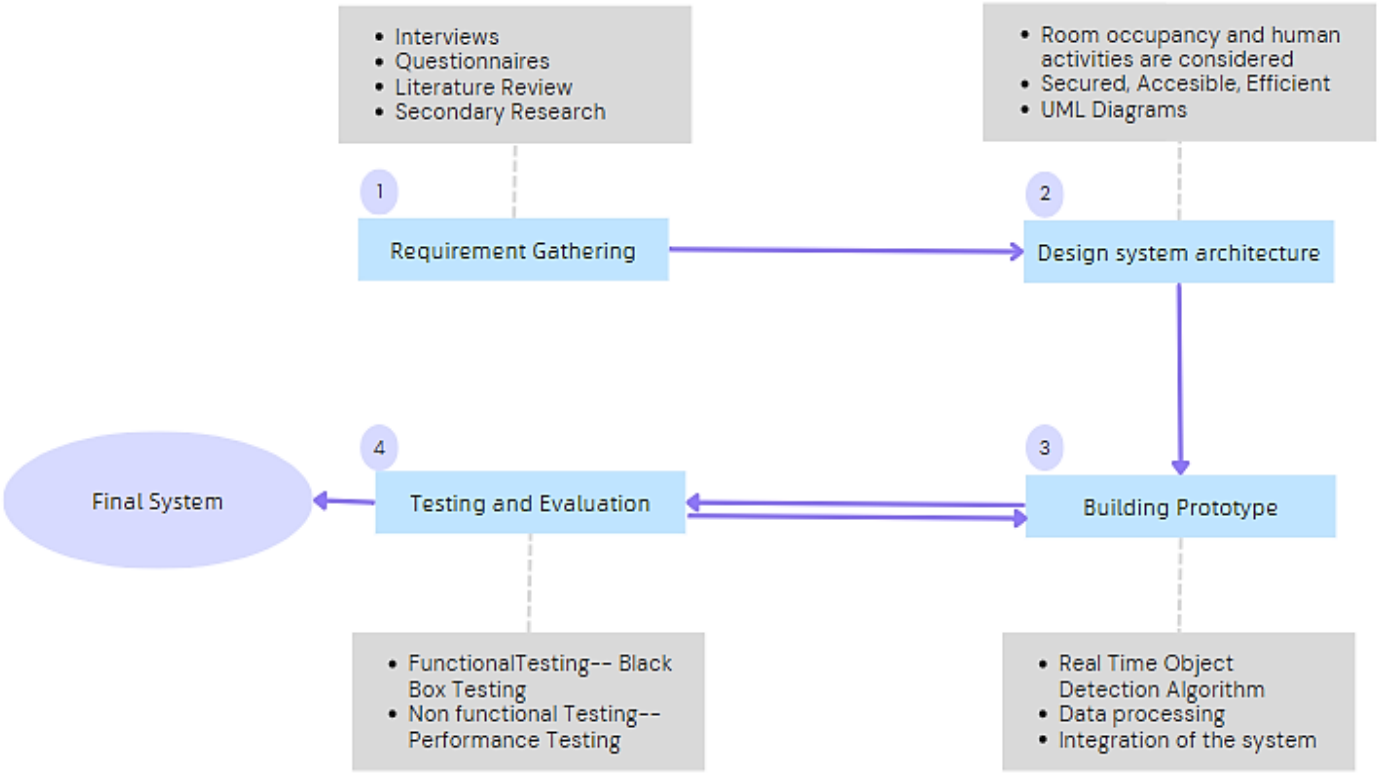


Fig. 1. SDLC prototype.

A. Phase 1: Requirements Gathering

Fig. 2 shows different preferences among respondents when adjusting air conditioner settings, preferences vary among respondents. A significant 76% prefer manual adjustments, while 39.6% use scheduled timers. Technological solutions are also employed: 14% use Wi-Fi remote control, 11% use smart thermostats, and 1% each use occupancy sensors and smartphone controls. Additionally, 2% utilize weather-based control, reflecting a blend of manual and tech-driven methods among the surveyed population.

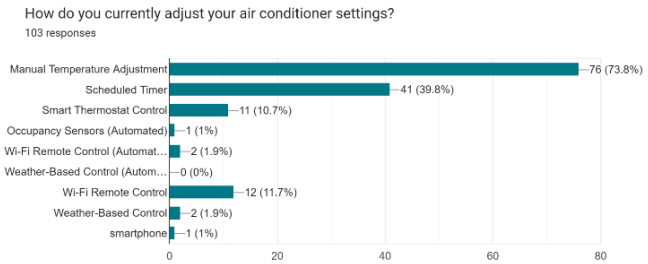


Fig. 2. Common air conditioner settings.

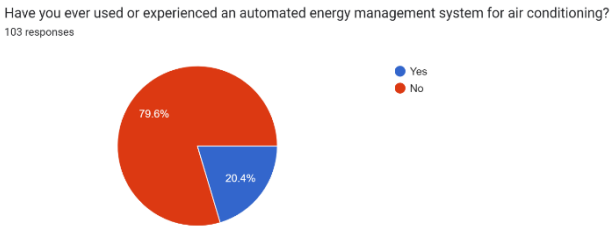


Fig. 3. Automated energy management system for air conditioning engagement report.

Fig. 3 shows that 79.6 per cent of respondents have not used or experienced an automated energy management system for air conditioning, while 20.4 per cent have reported having engaged with such automated systems. This distribution underscores that a significant majority of the surveyed population may not have had exposure to or utilized automated energy management systems in the context of air conditioning, highlighting potential opportunities for awareness and adoption of these technologies.

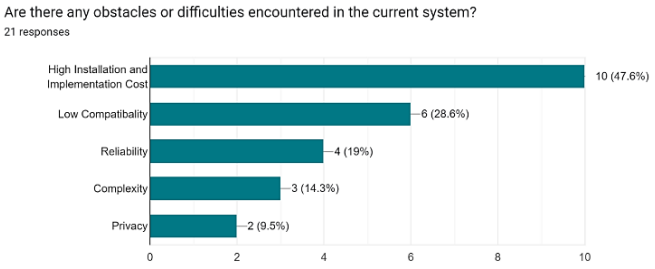


Fig. 4. Difficulties encountered in the current system report.

Fig. 4 shows that when asked about obstacles or difficulties encountered with the current air conditioning system, respondents identified several challenges. The most cited issue, chosen by 10 individuals, is the high installation and implementation cost, indicating a financial barrier that some face in adopting or maintaining their current system. Additionally, 6 respondents express concerns about low compatibility, highlighting challenges in integrating the system with other technologies or components. Four participants note reliability as an obstacle, suggesting issues with system dependability. Three respondents find complexity to be a challenge, indicating difficulties in navigating or understanding the operational intricacies of their air conditioning setup. Lastly, privacy is a concern for two individuals, suggesting a consideration of data security and personal information in the context of their air conditioning system. This diverse set of challenges provides valuable insights into the various hurdles faced by users in their experiences with current air conditioning systems.

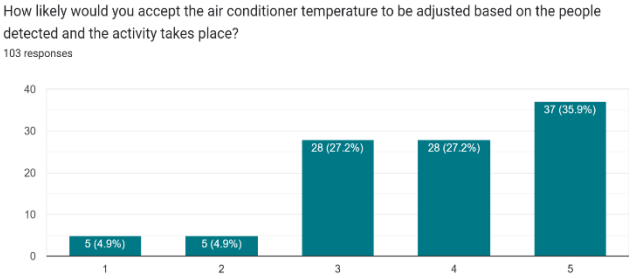


Fig. 5. Smart air conditioning acceptance rate.

Fig. 5 shows the data from respondents on their likelihood of accepting air conditioner temperature adjustments based on detected people and activities presents an interesting trend on the linear scale. A substantial 35.9 per cent expressed a high likelihood (rated 5) of accepting such adjustments, indicating a significant openness to embracing a system that adapts based on occupancy and activities. Furthermore, 27.2 per cent each chose the ratings of 3 and 4, suggesting a considerable portion with a moderate to high inclination. Meanwhile, 5 of the respondents each chose the ratings of 1 and 2, indicating a small minority expressing a lower likelihood. This data reflects a generally positive reception towards the idea of dynamically adjusting air conditioner temperature based on detected people and activities, with a notable majority leaning towards acceptance of such a system.

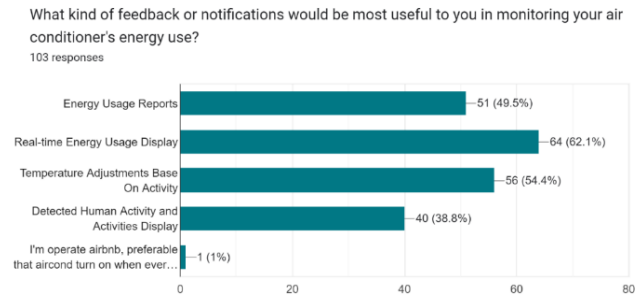


Fig. 6. Responses on preferred feedback for monitoring energy usage.

Fig. 6 shows in response to the inquiry about the most useful feedback or notifications for monitoring air conditioner energy use, most respondents, 62.1 per cent (64 individuals), express a preference for a real-time energy usage display. This indicates a strong interest in receiving immediate and transparent information about the energy consumption of their air conditioning system. Additionally, 54.4 percent (56 individuals) prioritize notifications related to temperature adjustments based on detected activities, highlighting a desire for a dynamic and responsive climate control system.

Moreover, 38.8 per cent (40 individuals) find value in notifications related to detected human activities and activities display, emphasizing the importance of understanding and being informed about the environmental conditions driving air conditioning usage. One additional response provides a unique perspective, as the respondent, operating an Airbnb, prefers the air conditioner to automatically turn on when human presence is detected and turn off when the guests leave, with the goal of optimizing energy savings. These insights collectively illustrate a strong demand for real-time information and dynamic control features that align with user preferences and enhance energy efficiency in the context of air conditioning systems.

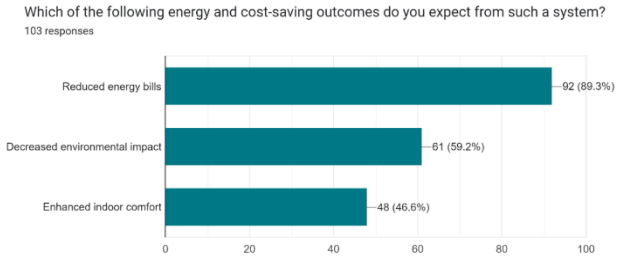


Fig. 7. Expectations of the system.

Fig. 7 shows the expectations regarding energy and cost-saving outcomes from the proposed system reveal a strong consensus among respondents. The most prominent expectation, chosen by 89.3 per cent of participants (92 individuals), is the anticipation of "Reduced energy bills." This aligns with a primary goal for many users, emphasizing the economic benefits associated with energy savings.

Additionally, 59.2 per cent of respondents (61 individuals) expressed the expectation of a "Decreased environmental impact," reflecting a heightened awareness and concern for

sustainability and environmental considerations. Furthermore, 46.6 per cent of participants (48 individuals) anticipate "Enhanced indoor comfort" as an outcome. This suggests that, beyond cost savings, respondents value the potential for improved comfort and well-being within their living or working spaces.

These expectations collectively underscore a strong interest in the practical and environmental benefits that respondents associate with the implementation of energy management systems in their air conditioning setups.

B. Phase 2: Design System Architecture

A Rich Picture is created by drawing the system with few instructions, often containing aspects such as architecture, processes, climate, people, issues stated by individuals, and conflicts. While lengthy writing is avoided, some Rich Pictures may include textual elements. The goal is to avoid mimicking organized diagrams such as Theory of Change maps while encouraging flexibility. Arrows can be used to express causal links in Rich Pictures. The analysis of a Rich Picture occurs during the process of drawing and discussing, making these diagrams a dynamic and flexible tool for understanding complex systems [9].

A Rich Picture shown in Fig. 8 is created by drawing the system with few instructions, often containing aspects such as architecture, processes, climate, people, issues stated by individuals, and conflicts. While lengthy writing is avoided, some Rich Pictures may include textual elements. The goal is to avoid mimicking organized diagrams such as Theory of Change maps while encouraging flexibility. Arrows can be used to express causal links in Rich Pictures. The analysis of a Rich Picture occurs during the process of drawing and discussing, making these diagrams a dynamic and flexible tool for understanding complex systems [9].

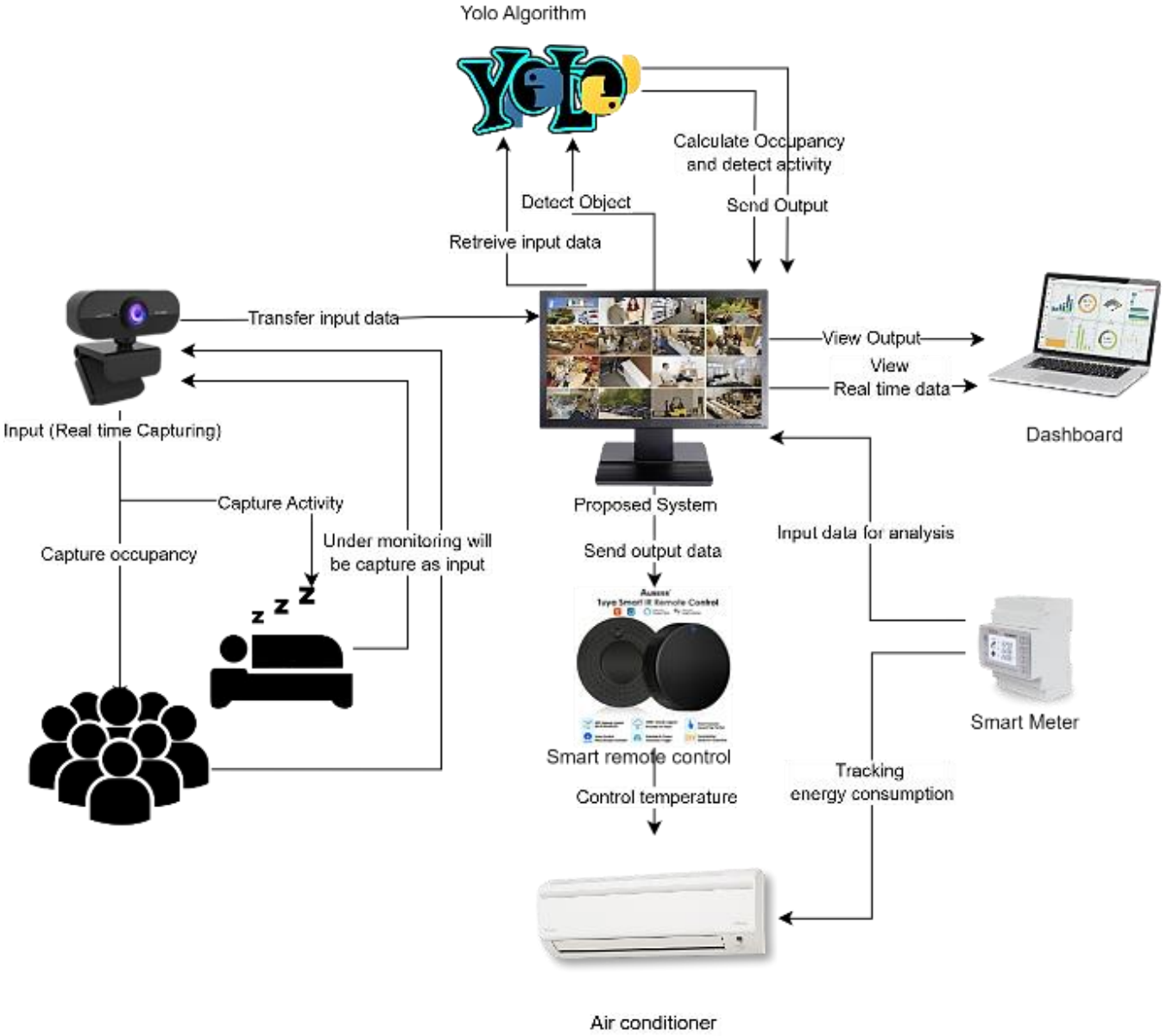


Fig. 8. Rich picture diagram.

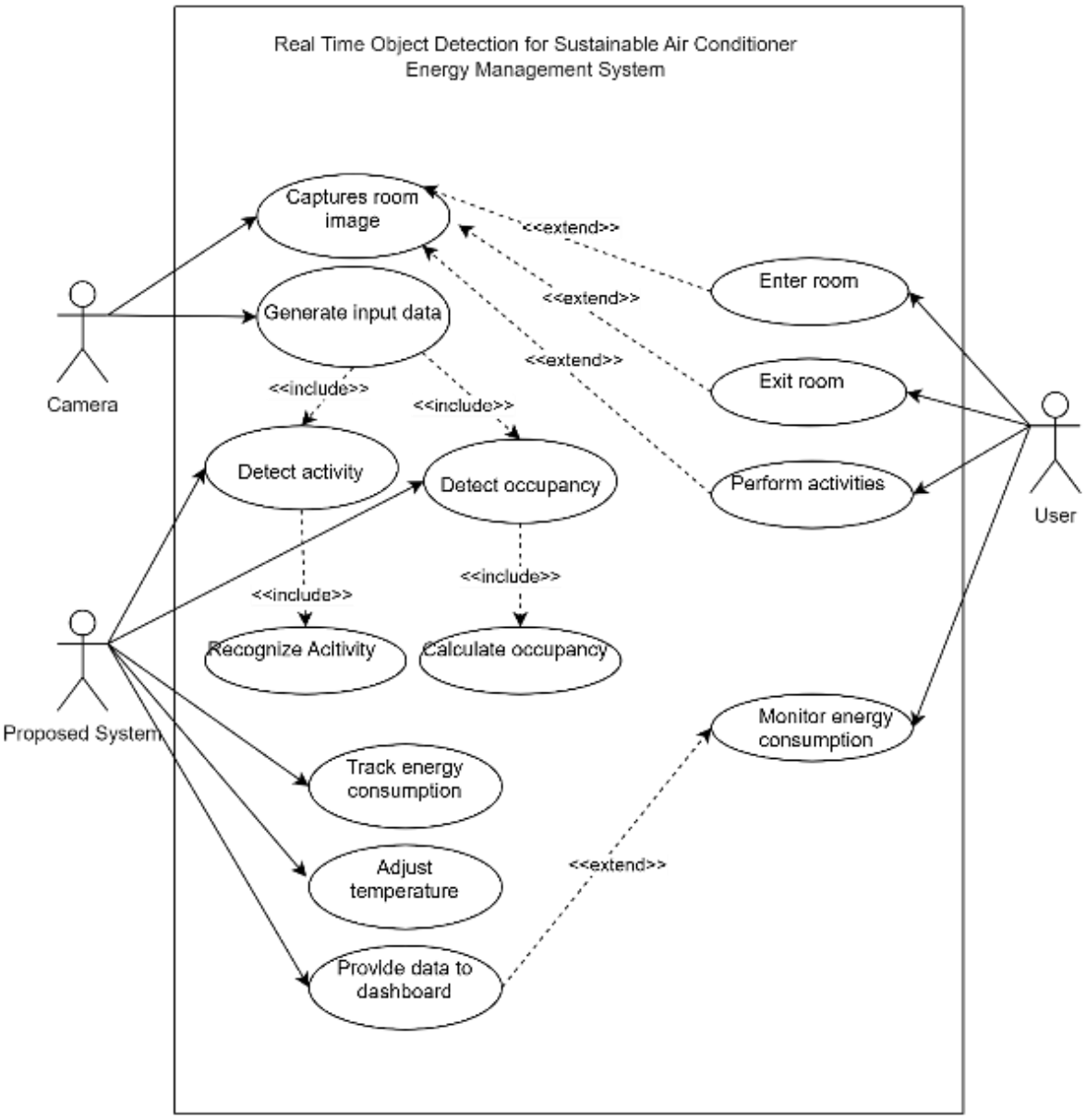


Fig. 9. Use case diagram of the proposed system.

Fig. 9 shows a use case diagram of the system. Use-case diagrams are essential UML tools for modelling system behaviour and capturing requirements. They provide a high-level overview of the system's functions, scope, and interactions with actors, illustrating what the system does without detailing underlying functions. These diagrams help define the context and requirements of the entire system or individual components. Typically created early in a project, they serve as valuable references throughout development, representing complex systems in one or multiple diagrams.

Fig. 10 shows the Class Diagram for the system comprises several distinct classes, each encapsulating specific functionalities within the overall architecture. The "Occupancy Detector" class is responsible for detecting and monitoring occupancy within the designated space, featuring attributes related to detection sensitivity and methods for processing occupancy data. The "Dashboard" class manages the graphical user interface, facilitating user interaction and real-time monitoring of system data. The "Account" class oversees user accounts, storing credentials and preferences, while the "Location" class represents spatial information tied to the physical context of the system. The "Smart Meter" class monitors and records energy consumption data, including parameters for readings. Lastly, the "Temperature Adjuster" class manages temperature settings, incorporating data related to control parameters, user preferences, and methods for adjusting temperature settings. Together, these classes define the static structure and responsibilities of key system components, providing a comprehensive overview of their interactions and collaborations. The Class Diagram serves as a visual guide to understanding the system's architecture and behaviour.

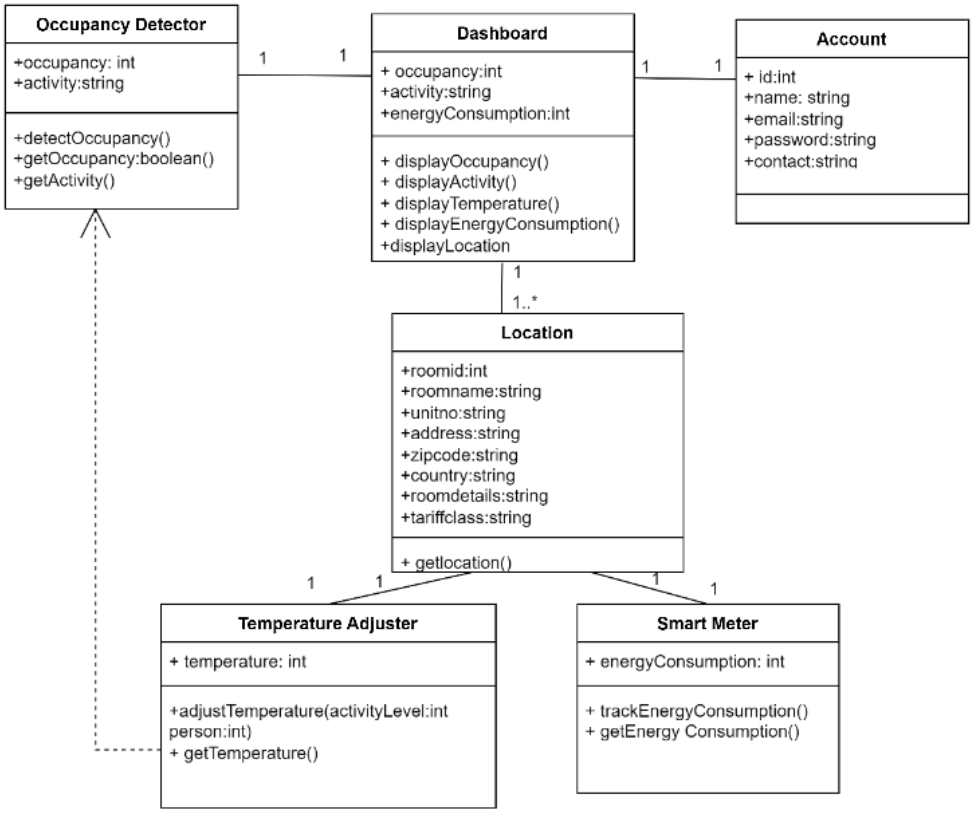


Fig. 10. Class diagram of the proposed system.

C. Phase 3: Building Prototype

Roboflow is employed to create dataset. The five classes in the dataset includes: sleeping, sitting, standing, moving, and exercising. There are a total of 2500 images and each class has an average of 250 pictures. All the images are annotated with the respective classes. 1855 (74%) images are used as train set, 448 images (18%) are used for valid set, and 197 images (8%) are used for testing purposes as shown in Fig. 11. Splitting the data into training, validation, and testing sets is important for developing an effective machine learning model.

Google Collaboratory is employed for the training of the YOLOv8 model. By leveraging Colab's infrastructure, accelerated processing capabilities are utilised, allowing for efficient training of the YOLOv8 model without the need for substantial local computational resources.

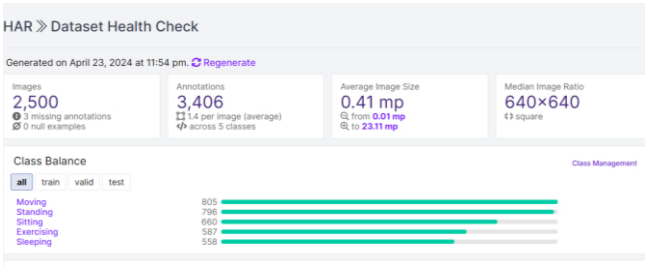


Fig. 11. Shows the health check of the dataset.

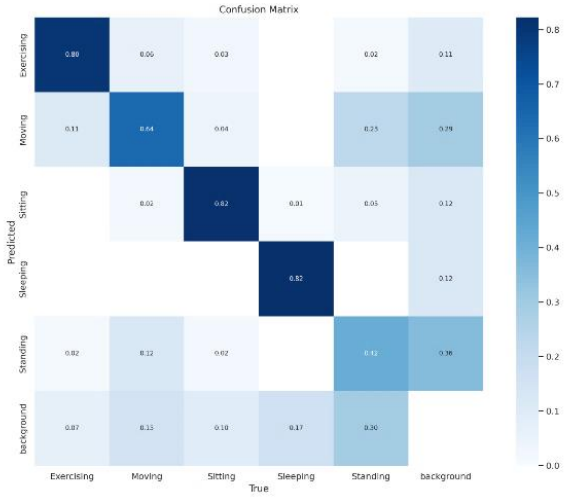


Fig. 12. YOLOv8m model confusion matrix.

As shown in Fig. 12, a high value on the diagonal 0.82 for "Sitting", 0.82 for "Sleeping" and 0.80 for "Exercising" means that the model accurately predicted that class most of the time. In contrast, lighter off-diagonal cells indicate misclassifications, such as "Standing" being incorrectly predicted as "Moving". The "background" class suggests that the model also tries to differentiate between specific activities and a lack of activity. This matrix can quickly reveal which classes are being confused by the model, guiding further improvements in its performance.

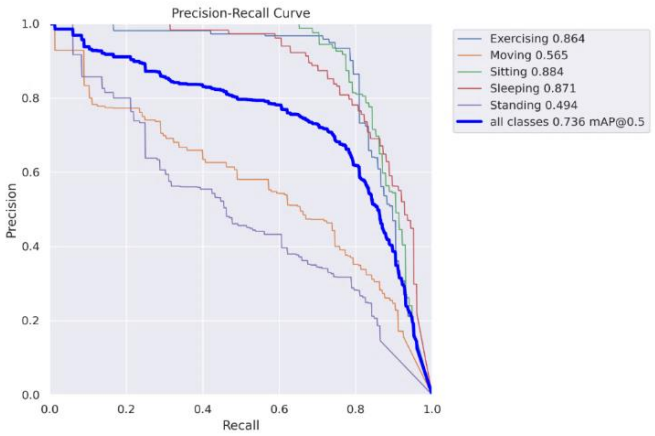


Fig. 13. YOLOv8m precision-recall curve.

The provided Precision-Recall curve in Fig. 13 evaluates a classification model's performance across five classes: “Exercising”, “Moving”, “Sitting”, “Sleeping”, and “Standing”. Precision generally decreases as Recall increases, reflecting the trade-off between accuracy and capturing all positive instances. “Sitting” exhibits the highest performance with an AUC (Area under curve) of 0.894, while “Standing” appears more challenging with a lower AUC of 0.494. The overall mAP at IoU (Intersection over Union) 0.5 is 0.736, indicating reasonable accuracy but room for enhancement, particularly in classes with less favorable Precision-Recall trade-offs.

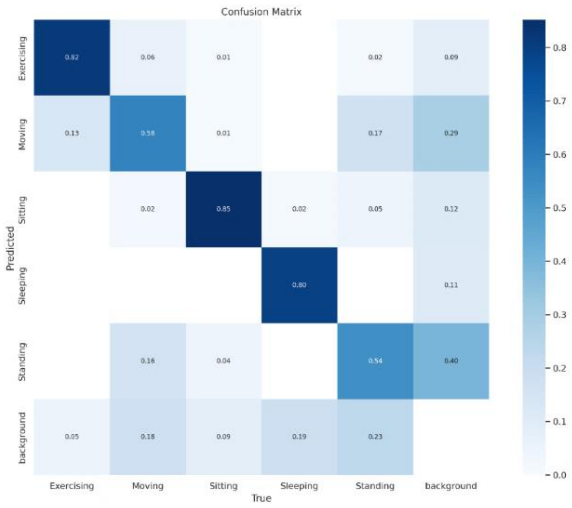


Fig. 14. YOLOv8x confusion matrix.

As shown in Fig. 14, a high value on the diagonal 0.85 for “Sitting”, 0.82 for “Exercising” and 0.80 for “Sleeping” means that the model accurately predicted that class most of the time. In contrast, lighter off-diagonal cells indicate misclassifications, such as “Standing” being incorrectly predicted as “Moving”. The “background” class suggests that the model also tries to differentiate between specific activities and a lack of activity. This matrix can quickly reveal which classes are being confused by the model, guiding further improvements in its performance.

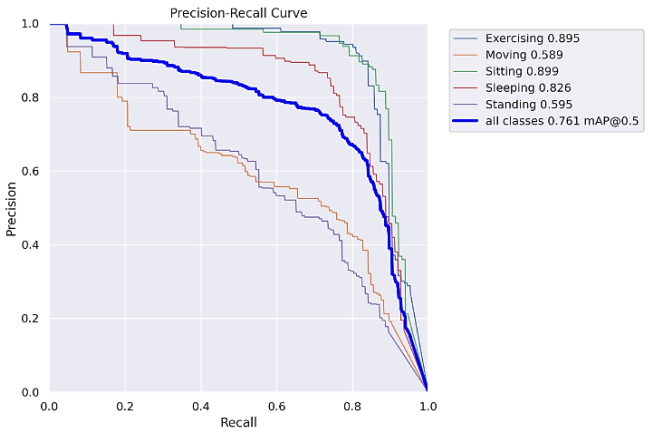


Fig. 15. YOLOv8x precision-recall curve.

The provided Precision-Recall curve in Fig. 15 evaluates a classification model's performance across five classes. Precision generally decreases as Recall increases, reflecting the trade-off between accuracy and capturing all positive instances. “Sitting” exhibits the highest performance with an AUC of 0.899, while “Moving” appears more challenging with a lower AUC of 0.589. The overall mAP at IoU 0.5 is 0.761, indicating reasonable accuracy but room for enhancement, particularly in classes with less favorable Precision-Recall trade-offs.

Based on these comparisons, the YOLOv8x model, which attained an accuracy of 76%, has been selected as the model for the Real-Time Object Detection system.

D. Phase 4: Testing and Evaluation

The testing plan encompasses both functional and non-functional aspects, including user registration, login, room management, profile editing, and system performance evaluations. Entry and exit criteria are established to ensure the proper setup and conclusion of the testing process, with a focus on achieving reliable and accurate system performance. Overall, the testing phase aims to validate the system's functionality and ensure its seamless operation in real-world scenarios, with detailed documentation provided to support the testing outcomes. The result from the usability testing shows that the system design is easy to use for the target user.

IV. DISCUSSION

The project successfully addressed the research question, “What software engineering method can be applied in the Air Conditioner Energy Management System to help reduce energy consumption?” by adopting YOLOv8 for real-time object detection and energy management. A comprehensive study of existing systems led to the development of a prototype integrating these functionalities.

To answer, “How to create a sustainable air conditioner energy management system that reduces energy consumption and contributes to a sustainable environment?” various tools were explored, system architecture was designed, and a user-friendly, efficient prototype was built as shown briefly in Fig. 16, Fig. 17, and Fig. 18.

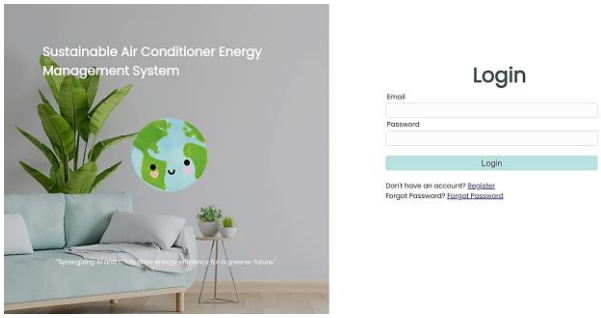


Fig. 16. Shows a screenshot of login page.

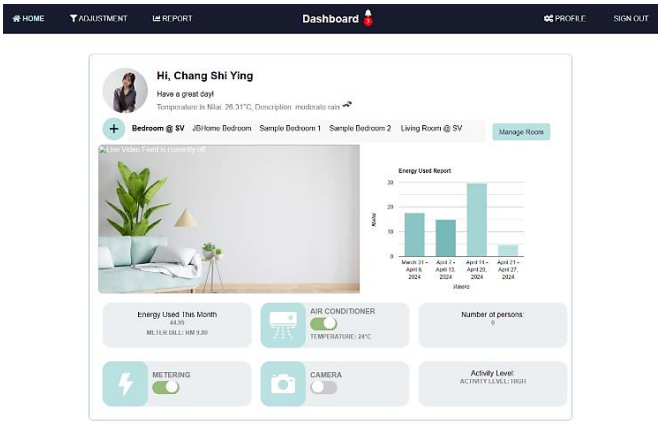


Fig. 17. Shows a screenshot of the home page (main dashboard).

Fig. 18. Shows a screenshot of the report page displaying real-world room smart meter data.

The project also evaluated the question, "How effective is the developed air conditioner energy management system in optimizing energy consumption in real-world scenarios?" by installing a smart meter to measure energy consumption before and after implementation. Testing validated the system's effectiveness and ensured seamless website functionality.

V. Results

The project successfully addressed the limitations of the existing system by adopting YOLOv8, offering significant improvements over the outdated YOLOv3 algorithm. YOLOv8 enabled more accurate real-time object detection and better tracking of room occupancy and activities, resulting in more precise temperature adjustments.

A key improvement lies in energy consumption management. The system optimized air conditioner energy usage by adjusting temperature settings based on real-time occupancy and activity data. This was a significant step forward compared to the previous system, which only counted the number of people in the room without accounting for specific activities. As a result, the new system reduced unnecessary energy consumption by avoiding overcooling or overheating the space.

The system also expanded accessibility by moving beyond Android-only compatibility, offering support for multiple operating systems and devices. Real-world testing confirmed these improvements. While no direct comparison with other systems was made due to project time limitations, the system enables users to monitor energy consumption through smart meter tracking, helping them avoid overuse. Website testing validated the seamless integration of the UI with smart home devices and common air conditioner models. The dashboard provides users with clear, actionable insights into their energy consumption, allowing for more informed energy management decisions.

However, limitations such as budget constraints impacting hardware performance, particularly in real-time object detection and dataset training, hindered optimal system performance. These challenges necessitated compromises in hardware capabilities and cloud platform usage, highlighting areas for improvement to enhance future project outcomes [10].

VI. Conclusion and Future Work

In conclusion, the developed system is deemed to effectively address the challenges outlined in the problem statement.

Enhancing the system's hardware infrastructure involves two key upgrades. First, investing in higher-performance GPUs, made possible with increased funding, accelerates model training and improves real-time object detection, boosting responsiveness and accuracy. Second, acquiring advanced cameras with superior low-light performance, faster capture speeds, and 360-degree adjustability enhances object detection, particularly in challenging lighting conditions and provides wider coverage.

Expanding the dataset and annotation process is crucial with increased resources. Additional time for dataset collection and annotation ensures a diverse range of images, improving the system's ability to accurately detect and classify objects. Including higher angle shots offers a broader perspective, further enhancing accuracy.

Strengthening system security is imperative for safeguarding user data. In addition to existing measures, implementing Secure Video Transmission, Camera Authentication, and Multi-Factor Authentication (MFA) bolsters defences against unauthorized access and cyber threats.

Integrating voice control or voice recognition based on user surveys enhances user experience and system usability. Voice commands for system functions streamline interactions, making the system more accessible and intuitive.

Future enhancements based on user feedback include refining the room selection dashboard with interactive buttons displaying meter data and editing options, integrating meter statistics directly into each room, allowing fixed temperature settings without camera detection, introducing usage limits with

notifications, and adding a leaderboard for energy-saving competition among users.

ACKNOWLEDGMENT

Acknowledgment is given to INTI International University for providing an invaluable platform for education and skill development. Special thanks are owed to the Faculty of Data Science and Information Technology for facilitating exposure to real-world challenges.

REFERENCES

[1] Rashed, A.H., Shah, A. The role of private sector in the implementation of sustainable development goals. Environ Dev Sustain 23, 2931–2948 (2021). https://doi.org/10.1007/s10668-020-00718-w.

[2] Omer, M.A.B. and Noguchi, T. (2019). A conceptual framework for understanding the contribution of building materials in the achievement of Sustainable Development Goals (SDGs). Sustainable Cities and Society, 52, p.101869. doi:https://doi.org/10.1016/j.scs.2019.101869.

[3] Sung, W.-T. and Hsiao, S.-J. (2020). The application of thermal comfort control based on Smart House System of IoT. Measurement, 149, p.106997. doi:https://doi.org/10.1016/j.measurement.2019.106997.

[4] Machorro-Cano, I., Alor-Hernández, G., Paredes-Valverde, M.A., Rodríguez-Mazahua, L., Sánchez-Cervantes, J.L. and Olmedo-Aguirre, J.O. (2020). HEMS-IoT: A Big Data and Machine Learning-Based Smart Home System for Energy Saving. Energies, 13(5), p.1097. doi:https://doi.org/10.3390/en13051097.

[5] Elsisi M, Tran M-Q, Mahmoud K, Lehtonen M, Darwish MMF. Deep Learning-Based Industry 4.0 and Internet of Things towards Effective Energy Management for Smart Buildings. Sensors. 2021; 21(4):1038. https://doi.org/10.3390/s21041038.

[6] Uddagiri Sirisha, S Praveen, Parvathaneni Naga Srinivasu, Paolo Barsocchi and Akash Kumar Bhoi (2023). Statistical Analysis of Design Aspects of Various YOLO-Based Deep Learning Models for Object Detection. International Journal of Computational Intelligence Systems, 16(1). doi:https://doi.org/10.1007/s44196-023-00302-w.

[7] Tien, P.W., Wei, S., Calautit, J.K., Darkwa, J. and Wood, C. (2020). A vision-based deep learning approach for the detection and prediction of occupancy heat emissions for demand-driven control solutions. Energy and Buildings, 226, p.110386. doi:https://doi.org/10.1016/j.enbuild.2020.110386.

[8] Wei, S., Tien, P.W., Calautit, J.K., Wu, Y. and Boukhanouf, R. (2020). Vision-based detection and prediction of equipment heat gains in commercial office buildings using a deep learning method. Applied Energy, 277, p.115506. doi:https://doi.org/10.1016/j.apenergy.2020.115506.

[9] Barbrook-Johnson, P. and Penn, A.S. (2022). Rich Pictures. Systems Mapping, pp.21–32. doi:https://doi.org/10.1007/978-3-031-01919-7_2.

[10] Leong, W. Y. (2023, August). Digital technology for ASEAN energy. In 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT) (pp. 1480-1486). IEEE.

The Review of Malaysia Digital Health Service Mobile Applications' Usability Design

Kah Hao Lim[1], Chia Yean Lim[2*], Anusha Achuthan[3], Chin Ernst Wong[4], Vina Phei Sean Tan[5]
School of Computer Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia[1, 2, 3]
Mooiko Software Sdn Bhd, 51/53, Jalan Zainul Abidin, 10450 Penang, Malaysia[4]
Department of Family Medicine-School of Medical Sciences-Health Campus, Universiti Sains Malaysia, 16150, Kelantan, Malaysia[5]

***Abstract*—Digital health services have become a trend and receive higher demand in Malaysia. However, the adoption of mobile applications to support the digital health service in the country remains low especially among older adults, contributing to low usability support of the mobile applications. This paper reviews the usability models and design factors that are relevant and applicable to support the digital health service mobile applications' design for older adults. Seven usability design factors such as efficiency, help and documentation, learnability, memorability, user-friendliness, need-base, and push-base were discovered to be most suitable to support older adult users. Subsequently, a review was conducted on the fulfilment of seven usability design factors in key Malaysian digital health service mobile applications. Findings showed that most applications supported high learnability and memorability but lacked support for another five usability factors. Lastly, a usability design framework to support the Malaysia digital health service mobile applications for older adult users would be proposed. A full exploratory study is the next step to validate the proposed framework.**

***Keywords*—Health system accessible; ISO/IEC9126; Nielsen usability model; older adults; usability**

I. INTRODUCTION

Smartphones play a vital role in our daily lives by providing connection and computational power for individuals and communities. Other than the common phone functions such as making calls and sending messages, the applications in current smartphones also support various activities such as information search, travel navigation, online shopping, personal health tracking, and remote health monitoring. Currently, about 98.4% of Internet users aged 16 to 64 in Malaysia own a smartphone compared to only 7% of the same group of people who owned feature phones [1]. Although older adults have switched from using feature phones to smartphones, the interface of the existing smartphones commonly presents small font sizes and rich content that targets mainly young people and not older adults [2]. Older adults may have different smartphone usage preferences compared to younger adults. Older adults prefer browsing messages and social media posts related to news, economics, and health. They will also download applications that provide support services related to their daily chores. If mobile applications have low visibility and low learnability, it may hinder older adults from accessing information and services through mobile applications and browsers.

Digital healthcare has been gaining traction in Malaysia for the last decade. It has further accelerated rapidly since the COVID-19 pandemic period. According to Digital 2024 Malaysia [1], about a third of the population or about 10.16 million Malaysians used digital health treatment and care applications. However, the average annual value per user for digital treatment and care was only recorded at 18.72 USD in the year 2023. The first reason for the low valuation of use is related to the limited digital healthcare services offered in Malaysia such as MySejahtera, DoctorOnCall, PruBSN Navigator, Doctor2U by BP Healthcare, and GetDoc. Secondly, digital health service mobile applications are difficult to use. According to Alharbi et al. [3] and Tajudeen et al. [4], older adults faced several issues while using health service mobile applications such as a lack of understanding about the features and the know-how to complete tasks in the application such as to book a service or call for emergency help.

While the life expectancy in Malaysia for Malaysia in 2024 is 73.8 and above the average life expectancy at birth of the global population at 71 years [5], there are more elderly people in Malaysia who need digital health services accessibility. The baby boomer generation may require more assistance in using mobile applications on smartphones. Unfortunately, the mobile applications' design is not user-friendly for older adults. A limited mobile application usability study has been conducted and similarly, a limited usability design framework has been proposed to support the digital healthcare application's development for older adults. Many significant usability factors such as intuitive interfaces, personalization, and push-based support were lacking in the current digital health service mobile applications to support the older adults' needs [6], [7], [8].

This review aims to study the usability models and explore the missing factors that could be used to support older adults in Malaysia in adopting digital applications for their health and well-being. This review aims to address three research questions which are: 1) Which usability models or frameworks are applicable for digital health service mobile applications? 2) Which common usability factors are adopted in current digital mobile application studies? and 3) What are the usability factors related to older adults that should be integrated into digital health service mobile applications? This review will conclude with a proposed usability design framework to

*Corresponding Author.

support the creation of healthcare mobile applications targeted at older adults.

This review paper has five main sections. Section I highlights the importance of supporting mobile application usability for older adults. Section II describes the methodology for conducting review activities. The findings of the review are explained in Section III. Follow on, a discussion of the findings of the review is written in Section IV before the conclusion and future work of this research are given in Section V.

II. METHODOLOGY

This review focuses on three key aspects. First, the existing technology acceptance models and usability frameworks are explored to determine the usability design factors. Second, the usability theory papers, existing conferences and journal publications related to usability evaluation of mobile applications that focus on digital health service applications are collected, synthesized, and summarized. Third, the existing digital healthcare mobile applications in Malaysia are explored and analyzed to determine the coverage of usability support in the applications. The search and review were conducted using Internet searches through the Google Scholar website and Google Play store dated from the year 2016 to 2024. The review duration covered the last 8 years of publications and mobile applications. In some cases, older references which are significant and impactful were also referred to and adopted.

III. FINDINGS OF THE REVIEW

The findings are presented based on the respective research question:

A. Findings on Research Question 1: Which Usability Frameworks are Applicable for Digital Health Service Mobile Applications?

Several theoretical frameworks and models were found to be relevant for digital health service mobile applications. The Technology Acceptance Model (TAM) proposed by Davis [9] is the foundational framework for understanding user acceptance of technology. It comprises two key determinants such as perceived usefulness and perceived ease of use which influence users' attitudes and actual intention to use technology (refer to Fig. 1). Current research still adopted this TAM to study the usability and acceptance of mobile applications such as mobile banking for Islamic banks [10], mobile e-wallets in Vietnam [11], and online food delivery applications [12].

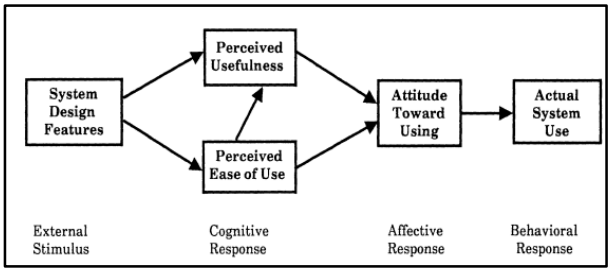


Fig. 1. Technology acceptance model (TAM) [9].

The TAM is incorporated into the Unified Theory of Acceptance and Use of Technology (UTAUT) model which was developed by Venkatesh et al. [13], together with the Theory of Reasoned Action (TRA) to provide a comprehensive understanding of the factors influencing the acceptance and usage of technology. The original UTAUT model as shown in Fig. 2, has four key determinants namely performance expectancy, effort expectancy, social influence, and facilitating conditions. Furthermore, this UTAUT model has been extended with usability attributes by Alshehri et al. [14] to cover more aspects of mobile application users' behavioural study. Besides the original four key determinants in the UTAUT model, six usability attributes namely system navigation, information quality, system learnability, visual design, system interactivity, and instructional assessment were included in the extended UTAUT model (refer to Fig. 3). The study could be moderated with demographic variables such as age, experience, gender, and willingness to use. These usability attributes were also popularly used and validated extensively in previous system evaluation studies [15], [16], [17], [18], [19]. This extended UTAUT model was adopted by Semiz & Semiz [20] who discovered that facilitating conditions attribute was the most significant factor, which affected the usage of mobile health applications.

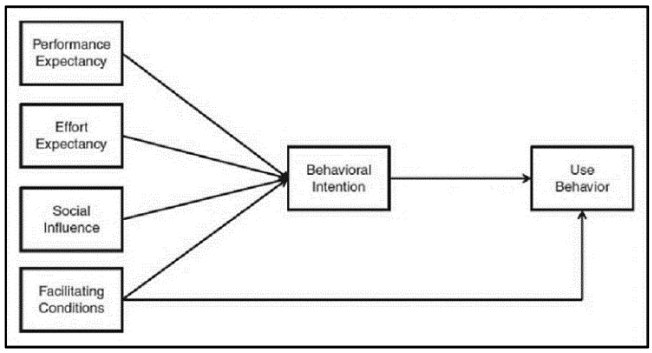


Fig. 2. UTAUT model [21].

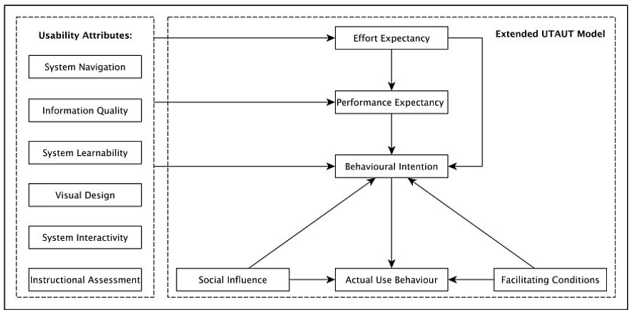


Fig. 3. UTAUT model with extended usability [14].

There are other usability theories and concepts proposed by other researchers. Nielsen proposed five quality components to support usability design [22]. The five quality components are learnability, efficiency, memorability, errors, and satisfaction. Much research was conducted and showed that the five quality components were significant in increasing the usability level of a product or system. The findings are summarized in Table I.

Nielsen also developed the Ten Principles of the Nielsen Usability Model [29]. Each principle provides insights and guidance for mobile applications, websites, and system designers to create user-friendly products to ensure that users can interact with the products efficiently and effectively. The

definition, applicability area for each of the principles, and the relevant research conducted are summarized in Table II.

TABLE I. SUMMARY OF RESEARCH FINDINGS FOR NIELSEN'S FIVE QUALITY COMPONENTS

Quality component	Definition of the Component	Supported research
Learnability	Refers to the ease of intended task accomplishment during the user's initial use of a software application	[23], [24], [25], [26], [27], [28]
Efficiency	Refers to the productivity level of using the system.	[23], [26], [27], [28]
Memorability	Refers to how easy to remember the system's functionalities and operations for users who return after having been on vacation or temporarily stopped using a program	[23], [26], [28]
Errors	Refers to any action that does not accomplish the desired goal	[23], [26], [27], [28]
Satisfaction	Refers to how pleasant the user is when using the system	[26], [27], [28]

TABLE II. OVERALL FINDINGS OF THE TEN PRINCIPLES OF NIELSEN USABILITY MODE

Principle	Definition	Applicability area	Supported research
Visibility of system	Refers to how well the system is conveyed to users	• Keep the users updated about what is going on, what is the latest status, where are the user located. • Support dynamic change of visuals and icons to highlight actions made. • Clarity of display messages.	[30]. [31]
Match between system and the real world	Refers to the natural mapping, which is easier for users to learn, use and remember how the interface and system works	• The UI element will reflect material objects similar to those objects from the real world. • Arrange the menu option and choose the color according to user's usual expectations. • The icons should convey information about their functions.	[32], [33], [34]
User control and freedom	Refers to the ability to allow users to rectify their mistake or backtrack their action such as back, cancel or undo.	• Provide browser back button. • The system "Exit" should be easy to find to let users in control when can they want to leave the system.	[34], [35]
Consistency and standards	Refers to the standardization of all the buttons and designs that exist in the system to lower the user's learning curve.	• Do not require the users to learn new interactions. • Follow established standards to make mobile interface design familiar to the users.	[31], [36], [37]
Error prevention	Refers to the effort to prevent users from making errors in the system.	• Eliminate error-prone conditions • Offer user confirmation before committing the action. • Use a placeholder and provide a range for data entry to eliminate data entry error.	[23], [34]
Recognition rather than recall	Refers to the effort to reduce the memory load of a user by making elements, actions, and options visible	• The information should be visible or easily retrieved when needed	[24], [38]
Flexibility and efficiency of use	Refers to the allowance of the user to use multiple interaction ways with the same result or task	• Allow the use of a shortcut key to reduce the number of clicks. • Define functional keys for the convenience of working with a system.	[24], [39]
Aesthetic and minimalist design	Refers to the design which consists of essential elements that support primary goals only.	• Hide rarely used or irrelevant page or module. • Encourage to use larger text and controls with lesser components in interface design.	[40], [41]
Help users recognise, diagnose, and recover from errors	Refers to the support given to help the users to recognise the errors and correct them.	• Provide meaningful error messages with plain language and not an error code. • Provide constructive suggestions to help the users to recover from the errors.	[34], [42]
Help and documentations	Refers to the availability of documentation that help users to better understand the system and know how to complete their tasks.	• Provide documents that contain a list of concrete steps that need to be carried out to complete a task. • Provide proactive help like tooltips, pop-up hints, and alert messages.	[24], [43]

Besides, the International Organization for Standardization (ISO) also created the ISO/IEC9126 matrix for information technology to evaluate their software quality in 1991 [44]. This matrix was expanded from year 2001 to 2004 to include the ISO quality model. Table III shows the five product quality characteristics in the ISO/IEC9126 matrix and their sub-characteristics. In the matrix, usability can be further explored in five aspects: understandability, learnability, operability, attractiveness, and usability compliance. Some of the sub-characteristics matched the quality components proposed by Nielsen. This matrix has been used by some researchers to investigate the system or software design efficacy towards the acceptance of users [45], [46].

TABLE III. CHARACTERISTICS OF ISO/IEC9126 MATRIX

Quality characteristic	Sub-characteristics
Functionality	• Suitability • Accuracy • Security • Interoperability • Functionality compliance
Reliability	• Maturity • Fault tolerance • Recoverability • Reliability compliance
Usability	• Understandability • Learnability • Operability • Attractiveness • Usability compliance
Efficiency	• Time behavior • Resource behavior • Efficiency compliance
Portability /Maintainability	• Analyzability • Changeability • Stability • Testability • Maintainability compliances

In conclusion, six related usability frameworks were studied and showed high applicability to be adopted in the digital health service mobile application design. Each framework proposed different variables. For example, TAM has two determinant variables; the UTAUT model has four key determinants; the extended UTAUT model has six usability attributes; Nielsen proposed five quality components as well as ten principles of the usability model; and the ISO/IEC9126 matrix comprises usability as product quality characteristic with five sub-characteristics. Each framework or model has been validated and used in various research. They are highly applicable in various types of mobile applications and web systems. Therefore, it is believed that these usability frameworks are significant and suitable for digital health service mobile application designs.

B. Findings on Research Question 2: Which Common Usability Factors are Adopted in Current Digital Mobile Application Studies?

A search was conducted on the Google Scholar website to explore the usability factors used in current digital mobile application studies. The search was conducted with keywords such as “mobile application”, “usability”, “user satisfaction”, and “user acceptance”. The search was filtered to include the recent 8 years’ research publications from 2016 to the current. The initial result returned 26,200 articles. The search results were later filtered to exclude duplicates, incomplete (pre-print, no full-text), and no usability factors as related to Research Question 1 finding, qualitative studies or studies with social science research frameworks. From the search conducted, 2006 hits were obtained and after applying filters, a total of 16 full-text articles were obtained. The mapping of 26 usability factors explored in the 16 current mobile application studies is displayed in Table IV.

The mapping in Table IV showed that efficiency, satisfaction, error, and effectiveness were the common usability factors used in existing studies. These variables were derived from Nielsen’s five quality components and ISO/IEC 9126 matrix. Other variables from TAM, UTAUT, and Nielsen’s ten principles of usability models such as learnability, ease of use, help and documentation, and visibility of system status were also adopted by the researchers. There are also some low-popularity variables such as need-base, good mapping, feedback, recognition rather than recall, consistency and standard, operability, and flexibility as displayed in Table IV.

TABLE IV. MAPPING OF USABILITY FACTORS TO CURRENT MOBILE APPLICATION STUDIES

Paper/ Variable	Need Base	Visibility of system status	Error	Consistency	Memorability	Recognition rather than recall	Ease of use	Effectiveness	Satisfaction	Learnability	Efficiency	Help and Documentation	Usefulness	Attitude or Likeability	Good conceptual Model	Good Mapping	Feedback	Security	Cognitive Load	Consistency and Standard	Privacy	Design	Accessibility	Aesthetics	Operability	Flexibility
[23]							✓	✓	✓		✓															
[24]		✓	✓			✓	✓					✓														
[26]			✓		✓				✓	✓	✓															
[27]			✓		✓				✓	✓	✓															
[28]			✓		✓				✓	✓	✓															
[31]	✓	✓																								
[34]												✓														
[37]												✓								✓						
[47]		✓	✓		✓		✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓							
[48]							✓	✓																		
[49]			✓					✓	✓	✓	✓										✓	✓				
[50]			✓		✓			✓	✓		✓		✓										✓	✓	✓	
[51]		✓	✓	✓	✓							✓											✓	✓		✓
[52]								✓	✓		✓															
[53]								✓	✓		✓															
[54]								✓	✓		✓															
Variable count	1	4	8	1	6	1	4	8	10	5	10	4	2	1	1	1	1	1	1	1	1	1	2	2	1	1

The 16 current mobile application studies focused mainly on applications with general users of all ages. These studies did not focus on the usability considerations for older adult users. For example, variables such as need-base, feedback, and accessibility that are assumed to be more relevant to older adult users were not explored adequately nor adopted in the current studies. It may be due to the nature of mobile applications that generally do not target older adult users, or researchers and developers may have neglected this group of users.

C. Findings on Research Question 3: What are the Usability Factors Related to Older Adults that Should be Adopted in the Digital Health Service Mobile Applications?

Based on the analysis and findings from Research Questions 1 and 2 above along with other relevant studies, seven usability factors were identified to support digital health service mobile applications that targeted older adult users. The factors were presented to a professional healthcare service provider and an industry-experienced user interface & user experience (UI&UX) developer to obtain their views on the usability factors' applicability to digital health service mobile application's design. Both experts agreed that this set of usability factors is relevant and suitable in Malaysia's context. The professional healthcare service provider also proposed two new variables (need-base and push-base) to be included, which do not belong to any existing usability model or framework, but are highly important for older adult users. Based on the input provided by both experts, the rationale for including the seven usability factors in digital health service mobile applications is derived and explained in Table V.

To examine whether the current studies have adopted the proposed usability factors for older adult users' support, the previous Table IV is filtered to show the studies that supported the seven proposed usability factors. The findings are displayed in Table VI. The filtering discovered that all 16 current studies included at least one of the usability factors that support older adult users, except for the push-base factor. A few current studies such as [26], [27], [28], [47] supported up to four usability factors for older adult users. The need-base and push-base factors are rarely or not supported by the current studies.

On the other hand, the researchers also reviewed the existing Malaysia-based digital health service mobile applications on the adoption of the seven proposed usability factors as presented in Table V. The review was conducted for six mobile applications namely MySejahtera [55], DoctorOnCall [56], BookDoc [57], PruBSN navigator [58], doctor2u By BP healthCare [59], and GetDoc [60]. The findings are shown in Table VII.

The digital health service mobile application that adopted the greatest number of usability factors for older adult support is doctor2u by BP Healthcare (Table VII). The mobile application only lacked the push-base component. The most downloaded health service mobile application in Malaysia – MySejahtera also fulfilled five usability factors for older adult support. Similarly, this mobile application also did not support push-base component, as well as the need-base component. The MySejahtera was created to support Malaysia COVID-19 pandemic healthcare activities such as health monitoring and vaccination. This mobile application was mandatory to be used by citizens of all ages in Malaysia during the movement control order period. Therefore, it is justifiable that the MySejahtera application's design has a high usability level to support older adult users. On the other hand, BookDoc, which is a personalised health and wellness mobile application for enhanced well-being is more dedicated to younger users and therefore only fulfilled one usability factor of high learnability related to older adult users.

In conclusion, the evaluation of various health service mobile applications revealed the shortcomings of the usability design in different aspects for older adult users. The help and documentation, learnability, and memorability design factors are well-adopted by current health service mobile applications. Meanwhile, the efficiency and user-friendliness design factors are only adopted by half of the current mobile applications. On the other hand, the need-base and push-base factors are only adopted by one mobile application reviewed. These findings indicate that most of the current health service mobile applications in Malaysia are not older adult-friendly in design. Currently, no Malaysian mobile applications for health adopted a similar design framework also. This implies a lack of standard usability design framework for health service mobile applications, especially to support older adult users.

TABLE V. RATIONALE FOR USABILITY FACTORS INCLUSION IN DIGITAL HEALTH SERVICE MOBILE APPLICATIONS

Variable	Rationale
Efficiency	Older adult users have lower attention span due to eye tiredness. The mobile application must support fast and timely task completion.
Help and documentation	Older adults may not be familiar with mobile application's functionalities and layout, a step-by-step documentation guide or responsive help in the mobile applications will help them to use the mobile application independently.
Learnability	Older adults may not have high digital literacy skills. The mobile application operations must be easy to learn and fast to pick up by the older adult users.
Memorability	Older adults may not remember as quickly or efficiently compared to younger adults. The mobile application should reduce the need for recall on how to use the mobile application's functions for older adults
User-friendliness	The mobile application should provide user-friendly components such as bigger fonts, suitable font colors and contrast, bigger buttons, and navigation controls to allow older adult users to view, click, and navigate the mobile application contents smoothly.
Need-base	Older adults may have special needs for features such as error prompting and correction guides, automated settings, and shortcuts that may not be required by younger age users.
Push-base	Older adults may have special needs for features such as service alerts and health reminders that may not be required by normal-age users.

TABLE VI. MAPPING OF CURRENT MOBILE APPLICATION STUDIES TO USABILITY FACTORS FOR ELDERLY SUPPORT

Paper	Efficiency	Help and documentation	Learnability	Memorability	User-Friendliness	Need - Base	Push-Base	Total fulfilment
[23]	✓	✕	✕	✕	✓	✕	✕	2
[24]	✕	✓	✕	✕	✕	✕	✕	1
[26]	✕	✓	✓	✓	✓	✕	✕	4
[27]	✓	✕	✓	✓	✓	✕	✕	4
[28]	✓	✕	✓	✓	✓	✕	✕	4
[31]	✕	✕	✕	✕	✕	✓	✕	1
[34]	✕	✓	✕	✕	✕	✕	✕	1
[37]	✕	✓	✕	✕	✕	✕	✕	1
[47]	✓	✕	✓	✓	✓	✕	✕	4
[49]	✓	✕	✓	✕	✓	✕	✕	3
[48]	✕	✕	✕	✕	✓	✕	✕	1
[50]	✓	✕	✕	✓	✓	✕	✕	3
[51]	✕	✓	✕	✓	✕	✕	✕	2
[52]	✓	✕	✕	✕	✓	✕	✕	2
[53]	✓	✕	✕	✕	✓	✕	✕	2
[54]	✓	✕	✕	✕	✓	✕	✕	2

TABLE VII. ANALYSIS OF MALAYSIA-BASED DIGITAL HEALTH SERVICE MOBILE APPLICATION IN USABILITY FACTORS' COVERAGE

Application	Efficiency	Help and documentation	Learnability	Memorability	User-Friendliness	Need - Base	Push-Base	Total fulfilment
MySejahtera [55]	YES	YES	YES	YES	YES	NO	NO	5
DoctorOnCall [56]	YES	YES	NO	YES	NO	NO	YES	4
BookDoc [57]	NO	NO	YES	NO	NO	NO	NO	1
PruBSN navigator [58]	NO	NO	YES	YES	NO	NO	NO	2
doctor2u By BP Healthcare [59]	YES	YES	YES	YES	YES	YES	NO	6
GetDoc [60]	NO	YES	YES	YES	YES	NO	NO	4

IV. DISCUSSION

This study discovered that the popular usability models adopted among the researchers are the Nielsen five quality components and ISO/IEC 9126 matrix. The researchers acknowledged the importance of having high usability coverage in the mobile application design to fulfil the specific users' needs and increase user satisfaction and experience. Research Questions 2 and 3 findings show different focuses on usability factors' adoption for general and health service mobile applications. It showed that the health service mobile application developers realised that the special-needs mobile application must have a unique set of usability factors to support the users. However, this awareness needs to be further strengthened to assist older adult users by considering more unique and related usability factors.

V. CONCLUSION

This study addressed three key research questions by using different approaches with reference to 43 significant research papers available in Google Scholar from the year 2016-2024. A theoretical study was conducted to identify and determine the relevant usability frameworks and models with the usability design factors to answer Research Question 1. As for Research Question 2, literature research was conducted in Google Scholar to explore the existing usability research for mobile applications in the past eight years. Studies adopted a different usability model for the mobile applications' design. It is observed that the Nielsen five quality components and ISO/IEC 9126 matrix are most adopted by researchers. Finally, to answer Research Question 3, seven usability factors were carefully selected and verified by the experts such as efficiency, help and documentation, learnability, memorability, user-friendliness, need-base, and push-base. Comparative analysis was also conducted on the existing 16 research papers and six Malaysia health service mobile applications. Overall, the fulfilment of usability factors to support older adult users was averagely low. This research identified the gap in determining the usability design framework that could be applied to mobile application design for older adults. This research is only limited to exploring the needs of older adults in the country of Malaysia as a pioneer study.

As the proposed usability design framework for older adult users is only preliminarily verified and agreed upon by two

experts, it is necessary to conduct a full exploratory study among the health service mobile application's older adult users to assess their agreement on the usability factors. The development of a health service mobile application prototype that covers the seven proposed usability factors as a proof-of-concept product for the older adult users' exploration is highly recommended. A quantitative analysis could then be conducted to measure the relationships between the seven usability factors and the acceptance or satisfaction level of health service mobile applications among older adult users. It is expected that once this future work is completed, a usability design framework to support health service mobile applications for elderly users could be proposed and adopted in Malaysia. This effort would increase Malaysian older adults' health system accessibility rate and digital healthcare value in Malaysia too.

ACKNOWLEDGMENT

This work was supported by the Universiti Sains Malaysia, Research University.

Team (RUTeam) Grant Scheme with Project No: 1001/PKOMP/8580021, Project.

Code: TE0035 (Reference No: 2022/0602).

REFERENCES

[1] S. Kemp, "Data reportal – global digital insights," Data Reportal. Available: https://datareportal.com/reports/digital-2024-malaysia. Accessed on 14 May 2024.

[2] Q. Li and Y. Luximon, "Older adults' use of mobile device: usability challenges while navigating various interfaces," Behaviour & Information Technology, vol. 39 (8), pp. 837–861, 2020, doi: 10.1080/0144929X.2019.1622786.

[3] R. A. Alharbi, F. T. Altayyari, F. S. Alamri, and S. A. Alharthi, "Pandemic-driven technology during COVID-19: Experiences of older adults," Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing, New York, NY, USA: ACM, 2021, pp. 5–9, doi: 10.1145/3462204.3481769.

[4] F. P. Tajudeen, N. Bahar, M. P. Tan, M. B. Peer Mustafa, N. I. Saedon, and J. Jesudass, "Understanding user requirements for a senior-friendly mobile health application," Geriatrics, vol. 7 (5), pp. 110, 2022, doi: 10.3390/geriatrics7050110.

[5] Countrymeters, "Malaysia population." Available: https://countrymeters.info/en/Malaysia. Accessed on 5 February 2024.

[6] C. Y. Wong, R. Ibrahim, T. A. Hamid, and E. I. Mansor, "Usability and design issues of smartphone user interface and mobile apps for older adults," Advances in Visual Informatics, vol. 11196, N. Abdullah, M. Badioze Zaman, and P. Robinson, Eds. Cham, Switzerland: Springer, 2018, pp. 93–104, doi: 10.1007/978-981-13-1628-9_9.

[7] I. Iancu and B. Iancu, "Designing mobile technology for elderly. A theoretical overview," Technological Forecasting and Social Change, vol. 155, pp. 119977, 2020, doi: 10.1016/j.techfore.2020.119977.

[8] Q. Li and Y. Luximon, "Older adults' use of mobile device: usability challenges while navigating various interfaces," Behaviour & Information Technology, vol. 39 (8), pp. 837–861, 2020, doi: 10.1080/0144929X.2019.1622786.

[9] F. D. Davis, "Technology acceptance model: TAM," Al-Suqri, MN, Al-Aufi, AS: Information Seeking Behavior and Technology Adoption, pp. 205–219, 1989.

[10] D. Suhartanto, D. Dean, T. A. T. Ismail, and R. Sundari, "Mobile banking adoption in Islamic banks," Journal of Islamic Marketing, vol. 11(6), pp. 1405–1418, 2019, doi: 10.1108/JIMA-05-2019-0096.

[11] T. To and T. H. M. Trinh, "Understanding behavioral intention to use mobile wallets in Vietnam: Extending the tam model with trust and enjoyment," Cogent Business & Management, vol. 8 (1), 2021, pp. 1891661, doi: 10.1080/23311975.2021.1891661.

[12] C. Y. Lim, S. L. Ang, Z.F. Chan, S. Li, "Food delivery applications with data mining methods: evaluating end-users' experience," Nanotechnology Perceptions, vol. 20 (2), pp. 107 – 115, 2024, doi: 10.62441/nano-ntp.v20iS2.9.

[13] V. Venkatesh, M. G. Morris, G. B. Davis, and F.D. Davis, "User acceptance of information technology: toward a unified view," MIS Quarterly, vol. 27 (3), pp. 425–478, 2003, doi: 10.2307/30036540.

[14] A. Alshehri, M. J. Rutter, and S. Smith, "The effects of UTAUT and usability qualities on students' use of learning management systems in Saudi tertiary education," Journal of Information Technology Education: Research, vol. 19, pp. 891–930, 2020, doi: 10.28945/4659.

[15] A. Alshehri, M. Rutter, and S. Smith, "Assessing the relative importance of an E-learning system's usability design characteristics based on students' preferences," European Journal of Educational Research, vol. 8 (3), pp. pp.839-855, 2019, doi: 10.21125/edulearn.2019.1304.

[16] M. M. Althobaiti and P. Mayhew, "Assessing the Usability of Learning Management System: User Experience Study," E-Education, and Online Training: Second International Conference, pp. 9–18, 2016, doi: 10.1007/978-3-319-28883-3_2.

[17] S. S. Binyamin, M. Rutter, and S. Smith, "Extending the Technology acceptance model to understand students' use of learning management systems in Saudi higher education," International Journal of Emerging Technologies in Learning (iJET), vol. 14 (3), pp. 4–21, 2019, doi: 10.3991/ijet.v14i03.9732.

[18] P. Zaharias and P. Koutsabasis, "Heuristic evaluation of e-learning courses: a comparative analysis of two e-learning heuristic sets," Campus-Wide Information Systems, vol. 29 (1), pp. 45–60, 2011, doi: 10.1108/10650741211192046.

[19] P. Zaharias and A. Poylymenakou, "Developing a usability evaluation method for e-learning applications: beyond functional usability," International Journal of Human-Computer Interaction, vol. 25 (1), pp. 75–98, 2009, doi: 10.1080/10447310802546716.

[20] B. Semiz and T. Semiz, "Examining consumer use of mobile health applications by the extended UTAUT model," Business & Management Studies: An International Journal, vol. 9 (1), pp. 267–281, 2021, doi: 10.15295/bmij.v9i1.1773.

[21] N. M. Suki and N. M. Suki, "Students' intention to use animation and storytelling: using the UTAUT model,", European Proceedings of Social and Behavioural Sciences, pp. 49–57, 2018, doi: 10.15405/epsbs.2018.05.5.

[22] J. Nielsen, "Usability Engineering," Morgan Kaufmann, 1994.

[23] S. Ajibola, E. T. Abiodun, and L. Goosen, "Development of a new model for the usability evaluation of M-commerce applications," Journal of Telecommunication, Electronic and Computer Engineering (JTEC), vol. 14 (3), pp. 1–9, 2022, doi: 10.54554/jtec.2022.14.03.001.

[24] H. Bouraghi, S. Rezayi, S. Amirazodi, E. Nabovati, and S. Saeedi, "Evaluating the usability of a national health information system with heuristic method," Journal of Education and Health Promotion, vol. 11 (1), pp. 182, 2022, doi: 10.4103/jehp.jehp_349_21.

[25] F. R. Jeddi, E. Nabovati, R. Bigham, and R. Farrahi, "Usability evaluation of a comprehensive national health information system: A heuristic evaluation," Informatics in Medicine Unlocked, vol. 19, pp. 100332, 2020, doi: 10.1016/j.imu.2020.100332.

[26] D. Yuniarto, and R. R. Marliana, "Evaluation of e-learning based on nau (nielsen's attributes of usability) according to students," J-Tin's-Jurnal Teknik Informatika, vol. 5 (1), 2021.

[27] N. B. Puspitasari, Y. Aulia, Z. F. Rosyada, "Usability evaluation of online transportation using Nielsen model," World Journal of Advanced Engineering Technology and Sciences, vol. 10 (2), pp. 1–9, 2023, doi: 10.30574/wjaets.2023.10.2.0283.

[28] J. P. Sibarani, "Usability and user satisfaction rate evaluation on e-learning application from student's perspective using Nielsen usability method," JURNAL INFOTEL, vol. 13 (3), pp. 120–127, 2021, doi: 10.20895/infotel.v13i3.673.

[29] J., Nielsen, "Ten Usability Heuristics", 2015.

[30] A. Harley, "Visibility of system status (usability heuristic #1)," Nielsen Norman Group. [Website], 2018, https://www.nngroup.com/articles/visibility-system-status/. Accessed on 11 December 2023.

[31] A. Nurhudatiana and J. Y. Seo, "An mhealth application redesign based on Nielsen's usability heuristics," in Proceedings of the 2020 The 6th International Conference on E-Business and Applications, New York, NY, USA: ACM, pp. 85–89, 2020, doi: 10.1145/3387263.3387267.

[32] A. Kaley, "Match between the system and the real world (usability heuristic #2)," Nielsen Norman Group. [Website], 2018, https://www.nngroup.com/articles/match-system-real-world/. Accessed on 11 December 2024.

[33] A. Afriansyah, W. Walhidayat, R. Novendra, L. Harefa, and S. Sutejo, "Usability testing on tracer study system using the heuristic evaluation method," Journal of Applied Engineering and Technological Science (JAETS), vol. 3 (2), pp. 178–184, 2022, doi: 10.37385/jaets.v3i2.807.

[34] F. R. Jeddi, E. Nabovati, R. Bigham, and R. Farrahi, "Usability evaluation of a comprehensive national health information system: A heuristic evaluation," Informatics in Medicine Unlocked, vol. 19, pp. 100332, 2020, doi: 10.1016/j.imu.2020.100332.

[35] M. Rosala, "User control and freedom (usability heuristic #3)," Nielsen Norman Group. [Website], 2020, https://www.nngroup.com/articles/user-control-and-freedom/. Accessed on 14 December 2023.

[36] R. Krause, "Maintain consistency and adhere to standards (usability heuristic #4)," Nielsen Norman Group. [Website]. https://www.nngroup.com/articles/consistency-and-standards/, Accessed on 1 December 2024.

[37] S. Saeedi, S. Amirazodi, S. Rezayi, and T. Khodaveisi, "Usability evaluation of two national COVID-19 registration systems: a heuristic evaluation," Avicenna Journal of Care and Health in Operating Room, vol. 1 (2), pp. 38–43, 2023, doi: 10.34172/ajchor.32.

[38] A. Joshi, D. M. P. Perin, C. Amadi, and K. Trout, "Evaluating the usability of an interactive, bi-lingual, touchscreen-enabled breastfeeding educational programme: application of Nielsen's heuristics," Journal of Innovation in Health Informatics, vol. 22 (2), pp. 265–274, 2015, doi: 10.14236/jhi.v22i2.71.

[39] P. Laubheimer, "Flexibility and efficiency of use (usability heuristic #7)," Nielsen Norman Group. [Website], 2020, https://www.nngroup.com/articles/flexibility-efficiency-heuristic/. Accessed on 16 December 2023.

[40] T. Fessenden, "Aesthetic and minimalist design (usability heuristic #8)," Nielsen Norman Group. [Website], 2021, https://www.nngroup.com/articles/aesthetic-minimalist-design/. Accessed on 9 November 2023.

[41] V. P. Cornet, C. N. Daley, P. Srinivas, and R. J. Holden, "User-centered evaluations with older adults: testing the usability of a mobile health system for heart failure self-management," Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 61 (1), pp. 6–10, 2017, doi: 10.1177/1541931213601497.

[42] N. Tim and E. Evan, "Error-message guidelines," Nielsen Norman Group. [Website], 2023, https://www.nngroup.com/articles/error-message-guidelines/. Accessed on 10 August 2023.

[43] J. Alita, "Help and documentation (usability heuristic #10)," Nielsen Norman Group. [Website], 2020, https://www.nngroup.com/articles/help-and-documentation/. Accessed on 11 February 2024.

[44] R. E Al-Qutaish, "Quality models in software engineering literature: an analytical and comparative study.," Journal of American Science, vol. 6 (3), pp. 166–175, 2010.

[45] Pan, T. C., and C. Y. Lim, "The exploration and analysis of Malaysia web 3.0 financial system's design factors with quantitative survey and clustering method", Nanotechnology Perceptions, vol. 20 (1), pp. 402 – 415, 2024, doi: 10.62441/nano-ntp.v20iS1.32.

[46] H. Xiao, and C. Y. Lim, "Exploration of factors affecting intention to use software testing tool in Jinan, Shandong, China with quantitative survey method," Nanotechnology Perceptions, vol. 20 (1), pp. 216-230, 2024. doi: https://doi.org/10.62441/nano-ntp.v20iS1.19.

[47] A. Alzahrani, V. Gay, and R. Alturki, "The evaluation of the usability in mobile applications," Proceedings of the 40th International Business Information Management Association (IBIMA), Seville, Spain, pp. 23–24, 2022.

[48] K. L. Hsieh, J. T. Fanning, W. A. Rogers, T. A. Wood, and J. J. Sosnoff, "A fall risk mhealth app for older adults: Development and usability study," JMIR Aging, vol. 1 (2), pp. e11569, 2018.

[49] M. Alshamari, "Usability Factors Assessment in Health Information System," Intelligent Information Management, vol. 8 (6), pp. 170–180, 2016, doi: 10.4236/iim.2016.86012.

[50] B. Maqbool and S. Herold, "Potential effectiveness and efficiency issues in usability evaluation within digital health: A systematic literature review," Journal of Systems and Software, vol. 208, pp. 111881, 2024, doi: 10.1016/j.jss.2023.111881.

[51] M. Mahmodi, M. Masomifard, N. KhatibZanjani, and M. Ahmadi, "Efficiency evaluation of E-learning courses at Payam Noor university based on learning usability criteria," Interdisciplinary Journal of Virtual Learning in Medical Sciences, vol. 11 (4), pp. 236–245, 2020.

[52] M. Georgsson and N. Staggers, "Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics," Journal of the American Medical Informatics Association, vol. 23 (1), pp. 5–11, 2016.

[53] K. R. Arthana, I. M. A. Pradnyana, and G. R. Dantes, "Usability testing on website wadaya based on ISO 9241-11," Journal of Physics: Conference Series, vol. 1165, p. 012012, 2019, doi: 10.1088/1742-6596/1165/1/012012.

[54] M. W. Iqbal, N. Ahmad, and S. K. Shahzad, "Usability evaluation of adaptive features in smartphones," Procedia Computer Science, vol. 112, pp. 2185–2194, 2017, doi: 10.1016/j.procs.2017.08.258.

[55] MAMPU *MySejahtera*. Unit Pemodenan Tadbiran Dan Perancangan Pengurusan Malaysia Government of Malaysia. Mobile app. Version 2.1.9. [Website], 2020, https://apps.apple.com/my/app/mysejahtera/id1504055630. Accessed on 14 January 2024.

[56] Health Digital Technologies Sdn Bhd. *DoctorOnCall - Online Pharmacy*. Health Digital Technologies Sdn Bhd. Mobile app. Version 4.20. [Website], 2022, https://apps.apple.com/my/app/doctoroncall-online-pharmacy/id1549810871. Accessed 14 January 2024.

[57] Provider Health4U Solutions Sdn. Bhd. *BookDoc - Go Activ Get Rewards*. Health4U Solutions Sdn. Bhd. Mobile app. Version 3.34.7. [Website], 2016, https://apps.apple.com/my/app/bookdoc-go-activ-get-rewards/id1037726744. Accessed on 14 January 2024.

[58] Prudential BSN Takaful Bhd. *PruBSN Navigator*. Prudential BSN Takaful Bhd. Mobile app. Version 1.15. [Website], 2014. https://apps.apple.com/my/app/prubsn-navigator/id760559895. Accessed on 14 January 2024.

[59] BP Diagnostic Centre Sdn. Bhd. *Doctor2U by BP Healthcare*. BP Healthcare Group. Mobile app. Version 5.6.1. [Website], 2015, https://apps.apple.com/my/app/doctor2u-by-bp-healthcare/id1001949029. Accessed on 14 January 2024.

[60] Jireh Group Pte Ltd. *GetDoc*. Jireh Group. Mobile app. Version 4.14.4. [Website], 2015, https://apps.apple.com/my/app/getdoc/id987814486. Accessed on 14 January 2024.

Badminton Tracking and Motion Evaluation Model Based on Faster RCNN and Improved VGG19

Jun Ou[1], Chao Fu[2]*, Yanyun Cao[3]
Physical Education Institute, Xinyu University, Xinyu 338000, Jiangxi, China[1, 2]
College of Physical Education and Health, Jiangxi Science and Technology Normal University, Nanchang 330000, China[3]

***Abstract*—Badminton, as a popular sport in the field of sports, has rich information on body motions and motion trajectories. Accurately identifying the swinging motions during badminton is of great significance for badminton education, promotion, and competition. Therefore, based on the framework of Faster R-CNN multi object tracking algorithm, a new badminton tracking and motion evaluation model is proposed by introducing a VGG19 network architecture and real-time multi person pose estimation algorithm for performance optimization. The experimental results showed that the new badminton tracking and motion evaluation model achieved an average processing speed of 31.02 frames per second for five bone points in the human head, shoulder, elbow, wrist, and neck. Its accuracy in detecting the highest percentage of correct key points for the head, shoulders, elbows, wrists, and neck reached 98.05%, 98.10%, 97.89%, 97.55%, and 98.26%, respectively. The minimum values of mean square error and mean absolute error were only 0.021 and 0.026. The highest resource consumption rate was only 6.85%, and the highest accuracy of motion evaluation was 97.71%. In addition, indoor and outdoor environments had almost no impact on the performance of the model. In summary, the study aims to improve the fast region convolutional neural network and apply it to badminton tracking and motion evaluation with higher effectiveness and recognition accuracy. This study aims to demonstrate a more effective approach for the development of badminton sports.**

Keywords*—Faster RCNN; VGG19; badminton; target tracking; motion evaluation

I. Introduction

With the popularity and popularization of badminton in international sports events, its training methods have gradually become diversified [1]. In order to better track targets in badminton sports scenes, prevent injuries caused by non-standard technical motions, and promote more standardized training, it is necessary to conduct in-depth discussions on badminton tracking and motion evaluation methods. Currently, common object tracking algorithms include Visual Object Tracking (VOT), Multiple Object Tracking (MOT), and Multi-Camera Multi Object Tracking (MCMOT) [2]. The Fast Region Convolutional Neural Networks (Faster RCNN) multi object tracking algorithm based on deep learning is currently the mainstream object tracking method in the field of motion detection [3]. Numerous researchers both domestically and internationally have explored the Faster RCNN multi object tracking algorithm. For instance, J. Meza et al. developed a Faster RCNN method on the basis of transfer learning to significantly improve public transportation and achieve real-time localization in highly occluded scenes. The method could effectively shorten travel time, improve road smoothness, thereby controlling the fleet and reducing congestion [4]. T. Shimizu et al. analyzed the difficulty of target tracking in open surgeries such as plastic surgery. A method for analyzing and evaluating open surgical videos was created by combining Faster RCNN localization, Residual Network 18 (ResNet-18), and Long Short Term Memory (LSTM) modules. The experimental results showed that this method successfully detected two different open surgeries. It was superior to the commonly used two baseline methods [5]. H. Li et al. aimed to optimize the management efficiency of urban sports public services, facilitate residents to exercise, and increase their happiness index. A smart target tracking model was proposed using the Faster RCNN algorithm. The experimental results showed that Faster RCNN had good accuracy and low average time. This model could guide different populations to fully utilize public service facilities, improve quality of life, and achieve good behavior in national sports [6]. X. Yin et al. designed an image object detection method on the basis of Faster RCNN to address the incomplete image feature extraction and low classification accuracy in existing image object detection algorithms. The experimental results showed that the average accuracy was 91.04%, which had good image target detection ability [7].

Although the Faster RCNN performs well in object tracking and detection, badminton is a complex sports scene that is prone to occlusion and light interference during the motion process [8]. Therefore, to improve the accuracy of badminton target tracking and detection, and reduce the loss, it is necessary to deepen the network hierarchy of the Faster RCNN. Visual Geometry Group19 (VGG19) is an architecture in deep neural networks. It adds more convolutional layers and parameters than other architectures, which can not only better extract image features but also better process more complex image data. It has been used in various visual detection fields [9, 10]. X. Wan et al. found that traditional machine vision algorithms couldn't successfully detect defects in various steel strips. Therefore, on the basis of fast image preprocessing algorithms and transfer learning theory, a complete improved VGG19 neural network strip defect detection process was proposed. The improved VGG19 had a recognition accuracy of 97.8%. Its performance in six types of defects outperformed the baseline VGG19 [11]. R. Mohan et al. proposed a VGG19 for diagnosing various lung diseases from chest CT images on the ground of customized medical image analysis and detection networks. The experimental results showed that in multi class classification tasks, the training accuracy and

testing accuracy of VGG19 performed excellently [12]. A. Faghihi et al. analyzed the skin lesion classification using Convolutional Neural Network (CNN) technology. A pre-trained neural network application transfer learning framework was constructed using VGG19. Compared with other methods, the classification accuracy of the method reached 98.18% [13]. To develop the non-invasive diagnostic method for Obstructive Sleep Apnea Hypopnea Syndrome (OSAHS) patients, L. Ding et al. proposed a pre-trained VGG19 and LSTM fusion model to classify the snoring sounds of simple snorers and OSAHS patients. The experimental results showed that the VGG19+LSTM had the highest classification accuracy of 99.31% for simple snorers snoring and OSAHS patients snoring [14].

In summary, current target tracking and detection technologies both domestically and internationally still face many challenges in dealing with occlusions and similar object interference in complex dynamic environments. Although various studies have attempted to enhance multi-object detection in images by integrating deep learning network models with the VGG19 architecture, as well as using new algorithms such as CNN and Faster RCNN, there is still a significant gap between the current detection performance and the expectation in practical applications. These gaps are mainly reflected in insufficient robustness, making it difficult to stably track targets in environments with high occlusion or similar object interference; real-time performance has not yet met the requirements of some application scenarios, especially in sports motion analysis that requires rapid response; limited generalization ability, with poor adaptability to data under different environments and conditions; and high consumption of computing resources, which restricts the application on devices with limited resources. Therefore, an innovative badminton tracking and motion evaluation model based on Faster RCNN and improved VGG19 is proposed in the study. By combining the powerful object detection and tracking capabilities of Faster RCNN with the VGG19 feature extraction, it can solve the existing challenges in badminton tracking and motion assessment technologies and further improve the accuracy of badminton tracking and motion assessment, thus providing a more efficient and accurate motion assessment solution for the field of badminton sports. This study is divided into five sections, first being the introduction. The second section introduces how the Faster RCNN target tracking algorithm is improved and how the optimized design model is established. The third section is performance testing of the new model. The fourth section is the discussion of the results. The last section is a summary of the paper.

II. Methods and Materials

In response to the existing problems in badminton tracking and motion evaluation, such as the challenge of dealing with severe occlusion and similar appearance interference in complex sports environments, this study first introduces the basic framework of Faster RCNN algorithm from the perspective of badminton target tracking. The VGG19 architecture is introduced and significantly improved. In addition, from the perspective of evaluating the motions of badminton players, the Faster RCNN-VGG19 target tracking algorithm is used as the framework foundation, taking the real-time multi person pose estimation algorithm (OpenPose) for further optimization. Through these improvements, a new comprehensive badminton tracking and motion evaluation method is ultimately proposed, aiming to improve the accuracy and real-time performance of badminton tracking and motion evaluation.

A. Construction of 3D Object Tracking Model Based on Faster RCNN and VGG19

In order to enable athletes to master the basic motions of badminton in a standardized manner and achieve precise and real-time motion evaluation, it is necessary to quickly and accurately detect and track moving targets. The Faster RCNN is a target detection algorithm in the RCNN series, which has strong target recognition capabilities [15]. It mainly contains two parts, namely the Region Proposal Network (RPN) and the target classification network based on target feature classification [16]. The RPN network and the object classification network share weight parameters, and the two networks are trained collaboratively, which can promote the network to have good robustness and accuracy. The Faster RCNN is displayed in Fig. 1.

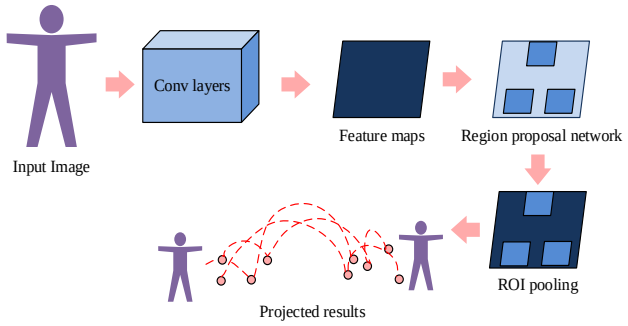


Fig. 1. Structure of the faster RCNN algorithm.

In Fig. 1, the Faster RCNN is mainly divided into three stages. Firstly, the motion image data is input into the network to obtain the corresponding feature image data. Secondly, the RPN is used to produce candidate boxes, mapping the candidate boxes generated by the RPN structure to the feature image data to obtain the relevant feature matrix. The obtained feature matrix is scaled to a size of 7×7 through the Region of Interest Pooling (ROI pooling) layer. Then, the 7×7 feature map is flattened and the final prediction result is obtained through fully connected layers. However, due to the limited appearance features of the small-sized shuttlecock, it is difficult to effectively distinguish the shuttlecock from similar small targets such as sneakers, light spots, and spectators' heads using the limited appearance features [17]. In addition, the traditional Faster RCNN uses the VGG16 framework, which cannot deepen the network hierarchy on the existing basis [18]. Therefore, in order to solve such problems, the study modifies the VGG16 framework in the Faster RCNN to the VGG19 framework. A new algorithm, namely the Faster RCNN-VGG19 object detection algorithm is proposed. The basic structure of the VGG19 framework is shown in Fig. 2.

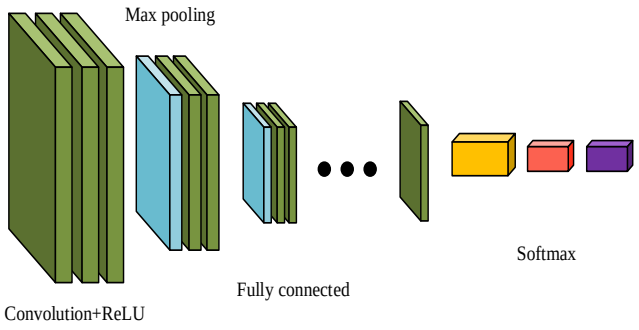


Fig. 2. Basic structure of the VGG19 framework.

In Fig. 2, the VGG19 framework mainly contains six parts: convolutional layer, batch normalization layer, ReLU activation function, maximum pooling layer, fully connected layer, and classifier [19]. The total depth of the network is 19 layers, including 16 CNN and 3 fully connected layers. The 16 CNN is further divided into 5 convolutional layers with varying numbers. The special structure of VGG19 can preserve all the features of the input image as much as possible, ensuring that the resolution of each layer's output and input is equal [20]. Batch normalization $y^{(q)}$ is shown in Eq. (1).

$$y^{(q)} = \gamma^{(q)} \frac{x^{(q)} - \mu^{(q)}}{\sqrt{(\sigma^{(q)})^2 + \varepsilon}} + \beta^{(q)} \tag{1}$$

In Eq. (1), $\gamma^{(q)}$ and $\beta^{(q)}$ represent learnable parameters. $x^{(q)}$ and $\mu^{(q)}$ represent the q-th dimensional input data and mean of the data, respectively. $\sigma^{(q)}$ represents the standard deviation. ε represents a number that prevents the denominator from being 0. Batch normalization can reduce the gradient vanishing and exploding, and accelerate the convergence speed of neural networks [21]. The ReLU is displayed in Eq. (2).

$$\mathrm{Re}\,LU(x) = \begin{cases} \max(x, 0), x \geq 0 \\ 0, x < 0 \end{cases} \tag{2}$$

In Eq. (2), x signifies the input data. When x is greater than or equal to 0, $\max(x, 0)$ is output. When x is less than 0, the output is 0. The ReLU (Rectified Linear Unit) activation function can perform a nonlinear transformation on the output of a neural network, thereby increasing the network's expressive and fitting capabilities. However, when the ReLU activation function encounters a situation where parameters need to be corrected during backpropagation, if the input is negative, the gradient becomes 0, which leads to an inability to adjust the parameters, resulting in the so-called "Dead ReLU" problem [22]. Therefore, the study introduces Leaky ReLU to address the DeadReLU of the ReLU in VGG19. The Leaky ReLU is shown in Eq. (3).

$$Leaky\,\mathrm{Re}\,LU(x) = \begin{cases} ax, x \geq 0 \\ x, x < 0 \end{cases} \tag{3}$$

In Eq. (3), a signifies the specified parameter, usually taking the smaller value. Leaky ReLU has a gradient even when the input is less than 0, and it possesses linear and non-saturating properties, which allows for fast convergence. It does not require exponential computations, making it computationally efficient and capable of addressing the issue of un-updatable weights in the standard ReLU activation function. The eigenvalue weight w_i is shown in Eq. (4).

$$w_i = \frac{e^{b_i}}{\sum_{j \in R} e^{b_j}} \tag{4}$$

In Eq. (4), b and R represent the feature map and local region, respectively. The eigenvalue weights ensure the transmission of important features, and during backpropagation, the features within the region will have a preset minimum gradient. Although the VGG19 framework can address the difficulties in tracking and recognizing small targets in the Faster-RCNN neural network algorithm, due to the effects of occlusion and lighting changes, there are still inevitable false positives and missed detections in the badminton detection results [23, 24]. Therefore, the study utilized the commonly employed triangulation algorithm from stereo vision matching to fuse the two-dimensional coordinates of the ball from multiple camera perspectives into three-dimensional coordinates, proposing a 3D target tracking model based on Faster RCNN and VGG19. The target tracking process framework of this model is shown in Fig. 3.

From Fig. 3, the target tracking process of the 3D target tracking model based on Faster RCNN and VGG19 is mainly divided into four stages: badminton 2D detection stage, badminton 2D tracking stage, badminton 3D coordinate fusion stage, and badminton 3D trajectory smoothing stage. The study first conducts badminton 2D detection in various camera perspectives based on the Faster RCNN-VGG19 object detection algorithm. Then, a 2D tracking algorithm based on Efficient Convolution Operators (ECO) is used to track the badminton balls in various camera perspectives. Secondly, the triangulation algorithm is used to effectively merge multiple 2D coordinates into one 3D coordinate. Finally, the Kalman filtering method is used to process and obtain smooth 3D badminton trajectories. The 3D coordinate point of badminton is shown in Eq. (5).

$$p_t = \frac{1}{N} \sum_{1 \leq i, j \leq n} (p_t^{ij} \left| e_t^{ij} < \tau \right.) \tag{5}$$

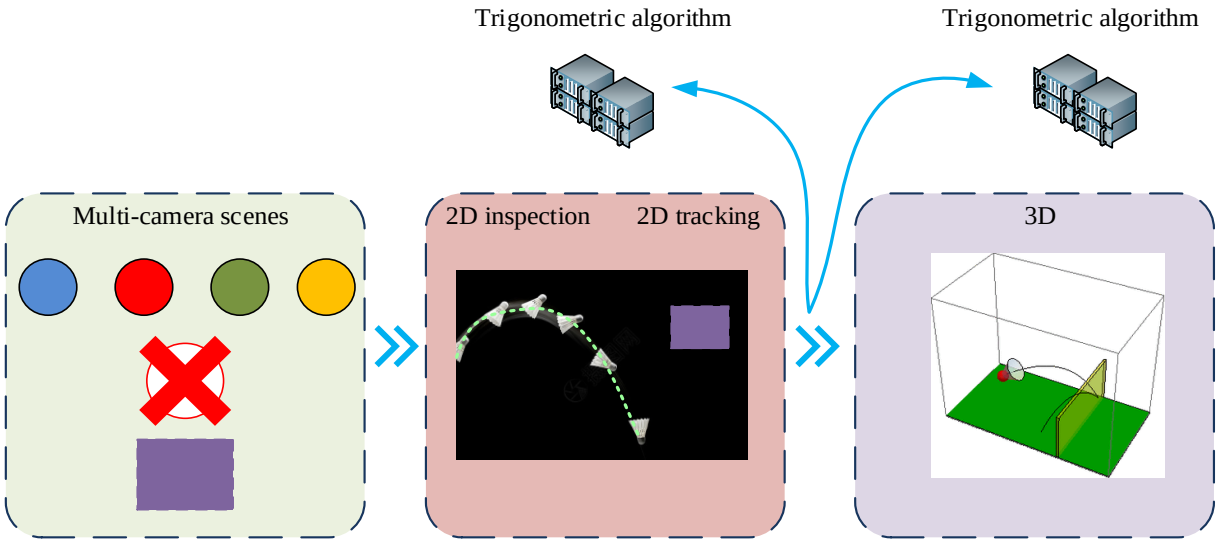


Fig. 3. Target tracking process framework for the proposed model.

In Eq. (5), p_t represents the three-dimensional coordinate fusion result of the badminton at time t. N signifies the number of camera matching pairs where the back projection error is less than the threshold τ. p_t^{ij} and e_t^{ij} represent the 3D coordinates and back projection errors calculated by the triangulation algorithm, respectively. The expression for maximizing probability $\hat{m}$ is shown in Eq. (6).

$$\hat{m} = \arg\max p(X_i^t = x_i^t, Y_i^t = Y_{j\in N(i)}^{t+1} \mid I^t) \tag{6}$$

In Eq. (6), X_i^t and x_i^t refer to whether there is badminton players or not, taking 0 or 1. Y_i^t and $Y_{j\in N(i)}^{t+1}$ represent the appearance characteristics of badminton players. The common evaluation metric for multi-object tracking algorithms is the Multiple Object Tracking Accuracy (MOTA). The study primarily takes into account three types of tracking errors — missed detections, false positives, and identity switches—for subsequent performance assessment. MOTA is shown in Eq. (7).

$$MOTA = 1 - \frac{\sum_t (c_1 \square fn_t + c_2 \square fp_t + c_3 \square idsw_t)}{\sum_t g_t} \tag{7}$$

In Eq. (7), c_1, c_2, and c_3 represent constants. g_t represents the true value. fn_t, fp_t and $idsw_t$ represent the number of missed detections, false detections, and number of identity exchanges, respectively.

B. Construction of Motion Evaluation Model Based on Faster RCNN and VGG19

After constructing a 3D object tracking model on the ground of Faster RCNN and VGG19, this study aims to address the various drawbacks of the badminton training system and attempt to optimize the model from the perspective of badminton motion evaluation. The first step in evaluating the motions of badminton players is to effectively obtain their posture information. Traditional methods for obtaining pose information often have drawbacks such as weak real-time performance, complex operation, and poor pose estimation performance [25]. The OpenPose can estimate the posture of the human body by analyzing key points in images or videos, identifying various parts of the body, and inferring the posture information. It can maintain accuracy even in complex environments [26, 27]. Therefore, OpenPose is introduced into the Faster RCNN-VGG19 object detection algorithm to analyze and process the posture information of badminton players. The basic structure of the Faster RCNN-VGG19-OpenPose pose estimation algorithm is shown in Fig. 4.

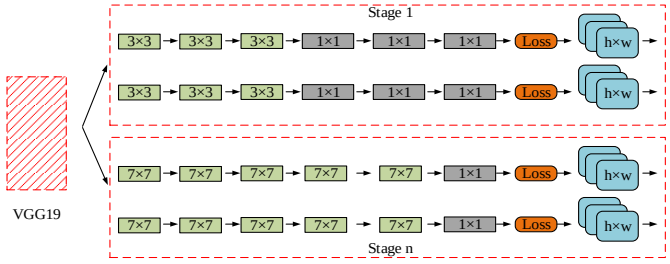


Fig. 4. The basic structure of OpenPose.

As shown in Fig. 4, the Faster RCNN-VGG19-OpenPose estimation algorithm is mainly divided into two parts, namely the limb confidence part and the site affinity vector field part. Firstly, the image is input into a parallel two branch Faster RCNN structure, and feature extraction is performed using the VGG19 architecture. Secondly, the feature maps of the image are obtained by clustering the bone points based on

initialization operation and greedy algorithm. Finally, the limb confidence prediction data and site affinity vector field prediction data are output through the limb confidence section and site affinity vector field. The predicted limb confidence data S^t and the predicted site affinity vector field data L^t are displayed in Eq. (8).

$$\begin{cases} S^t = \rho^t(F) \\ L^t = \varphi^t(F) \end{cases} \tag{8}$$

In Eq. (8), F represents the feature mapping. ρ^t and φ^t represent the CNN used for inference. In each subsequent stage, the original features and the two branch predictions generated in the previous stage are jointly input into the next stage, which can fully utilize the original features of the image to accurately predict the image in each stage. In addition, in order to successfully identify the categories of badminton swing motions, the research also cascades a fast Support Vector Machine (SVM) classifier based on the Decision Tree (DT) at the backend of OpenPose. The DT-SVM classifier can decompose nonlinear optimization problems into multiple linear SVM problems for solution, making the method simple and easy to implement. The DT structure diagram for badminton swing is shown in Fig. 5.

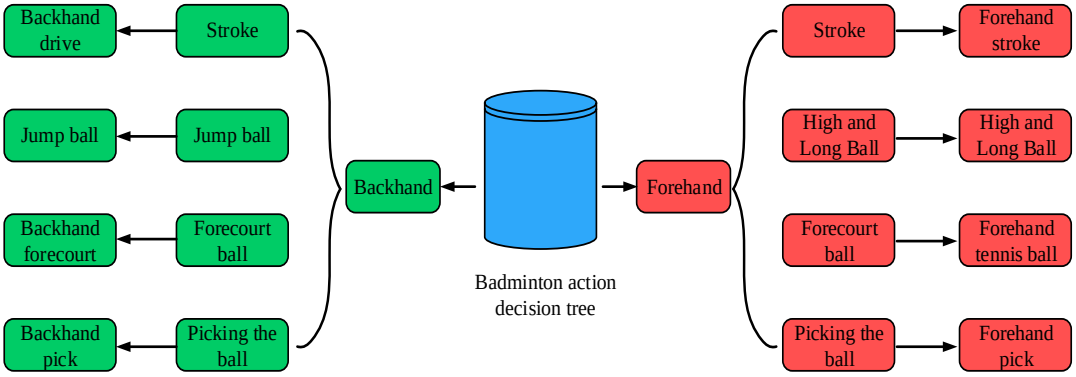


Fig. 5. Badminton swing decision tree.

From Fig. 5, the DT divides badminton swing motions into eight categories, namely high and long ball, Forehand stroke, forehand pick, Forehand tennis ball, jump ball, backhand drive, backhand pick, and backhand forecourt [28]. Firstly, the eight types of swing motions are combined to form a DT. During the classification process, each category needs to be selected before entering the next level, and unselected subtrees are deleted. This can effectively reduce the samples that need to be classified in the next step. The search step is repeated until the leaf node is reached. The final output result is obtained. However, the number of players in badminton training is generally large, and it is necessary to associate the results of the motion assessment with the target personnel; otherwise, the obtained sports assessment results will become meaningless [29]. Considering such situations, the study utilized the Particle Filter (PF) algorithm to deal with the non-Gaussianity of probability distributions in the tracking environment. The PF is shown in Eq. (9).

$$\begin{cases} x[t] = Ax[t-1] + v \\ y[t] = Cx[t] \end{cases} \tag{9}$$

In Eq. (9), $x[t]$ and $y[t]$ signify the coordinate values of the human skeletal neck. A represents the state transition matrix. v and C represent position random shift variables and diagonal matrices, respectively. The likelihood p^m of particles is shown in Eq. (10).

$$p^m = \frac{1}{\sqrt{2\pi\alpha^2}} \exp\left(-\frac{(d^m)^2}{2\alpha^2}\right) \tag{10}$$

In Eq. (10), d^m and α represent the degree difference and adjustable parameters, respectively. There are generally three annotation methods for human joint points: target instance annotation, target key point annotation, and image understanding annotation [30]. The key points of the human body during badminton motion are annotated based on target key point annotation. The annotation of key points in the human body is displayed in Fig. 6.

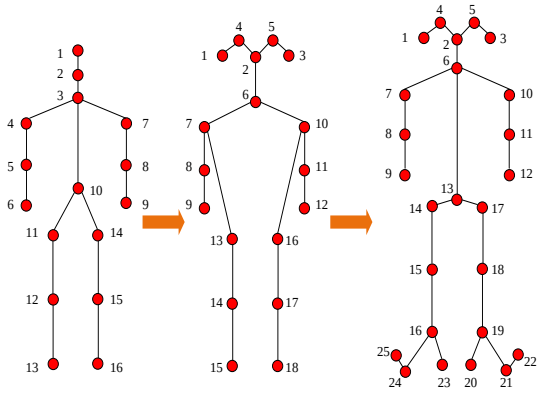


Fig. 6. Human body part key point labeling.

According to Fig. 6, there are three types of annotations for key points in the human body during badminton, namely 16 key points, 18 key points, and 25 key points. Among them, 18 key points increase facial key points compared with 16 key points, and 25 key points increase foot key points compared with 18 key points. The description of these three types of joint points can represent the skeletal information of the human body in detail, but these three methods will bring considerable computational complexity. Therefore, the study adjusts it by removing useless joints such as eyes and ears. The number of described joints is adjusted to 14. The distance $dis(x_i, x_j)$ between two joint points is shown in Eq. (11).

$$dis(x_i, x_j) = \sqrt{(x_i - x_j)^T M (x_i - x_j)} \tag{11}$$

In Eq. (11), x_i and x_j represent the positions of two joint points. M represents a symmetric positive semi-definite matrix. The rough pose representation metric is shown in Eq. (12).

$$E_{coarse}(W, B^*) = \left\| (x_n - x_n B^*)^T W_n \right\|_F^2 \tag{12}$$

In Eq. (12), B^* represents the coefficient matrix. x_n represents rough posture data, W then it represents the rank matrix of the entire column. Based on the above improvements, a motion evaluation model based on Faster RCNN and VGG19 is ultimately proposed. The motion evaluation process of this model is shown in Fig. 7.

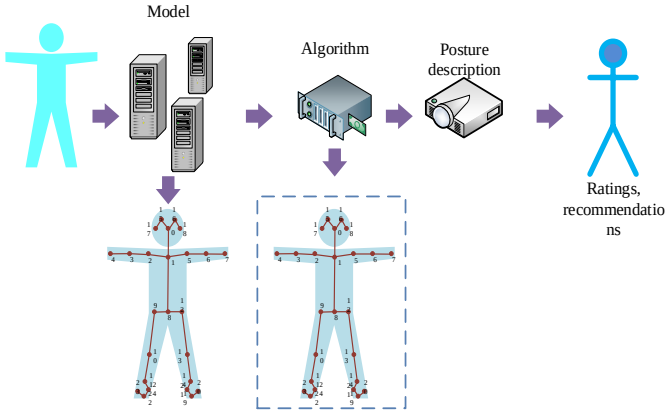


Fig. 7. The motion evaluation process of the proposed model.

From Fig. 7, the motion evaluation process mainly consists of four steps, namely the 3D target tracking model based on Faster RCNN and VGG19, posture estimation, granular posture description and scoring, and proposing motion suggestions. Firstly, a three-dimensional target tracking model based on Faster RCNN and VGG19 is used to quickly and accurately obtain the three-dimensional coordinates of badminton. Then, the position information of all human bones in the image is determined, and the prior standard parameters for each motion are obtained. Next, the posture image is described. Finally, the motion standard level is evaluated and motion suggestions are proposed.

III. Results

A. Faster RCNN-VGG19-OpenPose Algorithm Performance Testing

To verify the performance of the improved Faster RCNN-VGG19-OpenPose algorithm, a suitable experimental environment is established. The CPU is set to Intel Core i7 with a base frequency of 4.2Hz. The GPU is set to NVIDIA GeForce RTX 1660s, with 16GB of graphics memory and 16GB of memory. The Windows 10 is the operating system, Python as the algorithmic language. MPII and COCO datasets are used as the test data sources. The MPII dataset is a database of human body postures, containing approximately 25000 images and 40000 human body node information from different postures, and covering over 410 activities, all of which are sourced from YouTube videos. The COCO dataset is a dataset provided by the Microsoft team that can be used for image recognition, including various motion scenes. The study divided these datasets into training and testing sets in a ratio of 6:4, set the initial learning rate to 0.001, processed 100 frames per batch, and set the weight coefficient to 1. The study first conducts ablation tests on the Faster RCNN-VGG19-OpenPose algorithm using detection accuracy as an indicator. The test results are shown in Fig. 8.

Fig. 8(a) shows the ablation test results of each module of the Faster RCNN-VGG19-OpenPose algorithm on the training set. Fig. 8(b) shows the ablation test results of each module of the Faster RCNN-VGG19-OpenPose algorithm on the testing set. As shown in Fig. 8, with the continuous increase of iterations, the detection accuracy of each module showed an upward trend. The optimal detection accuracy of the basic Faster RCNN was 73.29%. After introducing the VGG19 architecture for optimization, the overall performance of the module improved by about 11%. The Faster RCNN-VGG19-OpenPose algorithm proposed in the study had the best performance, with the highest detection accuracy of 94.69% in the training set and a minimum of 298 iterations. The highest detection accuracy in the testing set was 93.08%, with a minimum of 243 iterations. From this, each module of the proposed algorithm has a positive effect on the overall performance, and the effect is significant. In addition, the study introduces popular target tracking algorithms for comparison, such as VOT algorithm, MOT algorithm, and MCMOT algorithm. A comparison test is conducted using tracking accuracy as an indicator, with a threshold of 200cm, as displayed in Fig. 9.

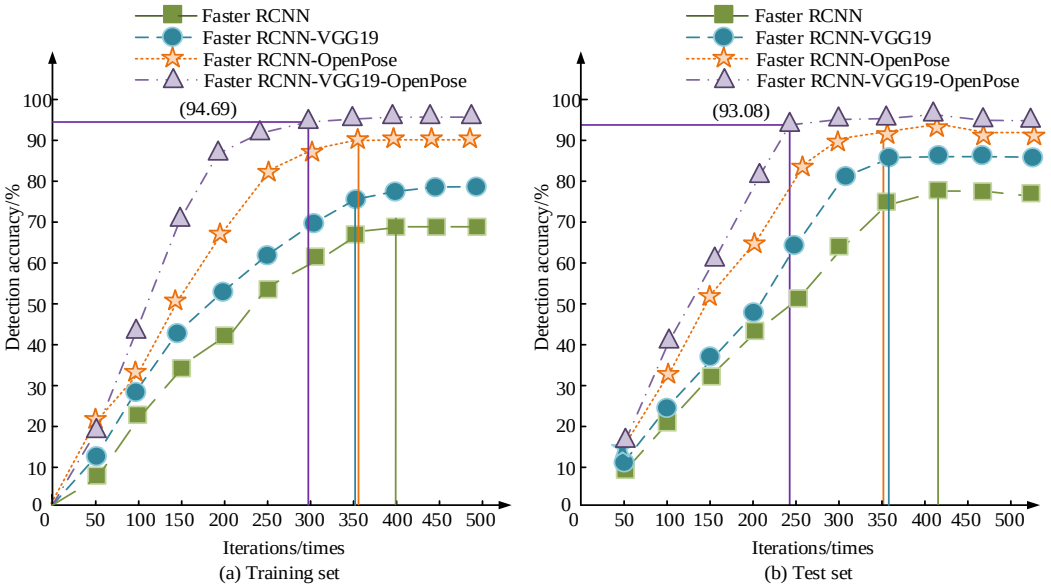


Fig. 8. Ablation test results of target tracking module with different datasets.

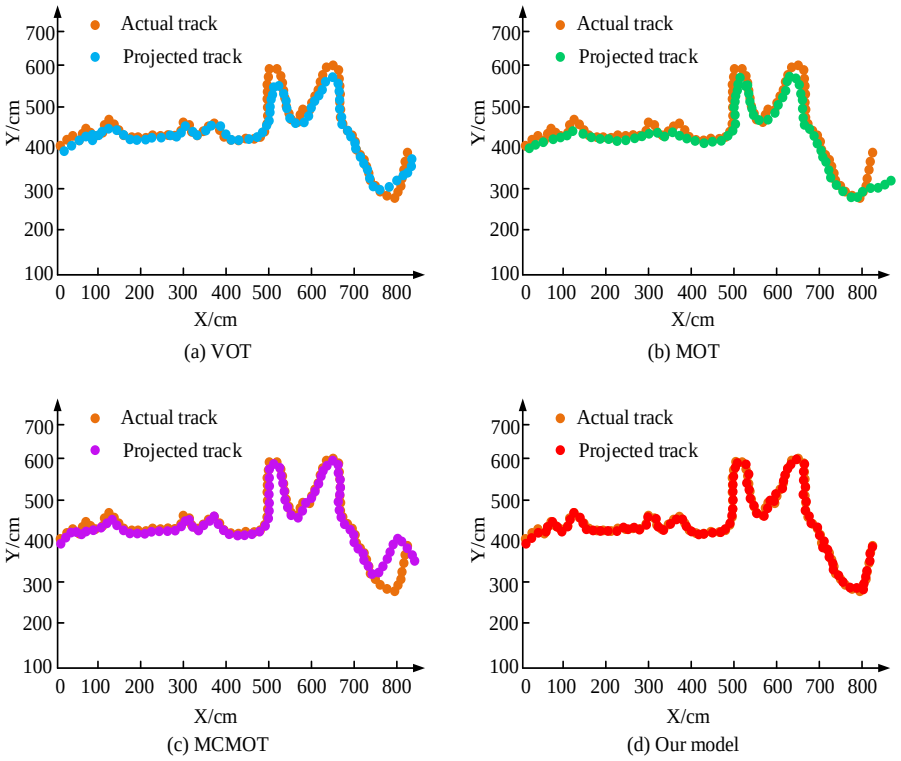


Fig. 9. Tracking accuracy of different target tracking algorithms.

Fig. 9(a) presents the comparison results of the VOT algorithm. Fig. 9(b) presents the MOT algorithm. Fig. 9(c) presents the MCMOT algorithm. Fig. 9(d) presents the Faster RCNN-VGG19-OpenPose. From Fig. 9, the Faster RCNN-VGG19-OpenPose algorithm had the highest overlap between the real trajectory and the tracked trajectory, followed by the MCMOT algorithm, and the worst overlap between the MOT algorithm and the VOT algorithm. This indicates that the proposed algorithm has strong robustness against common missed and false detections in detection. The VGG19 architecture deepens the hierarchy of Faster RCNN, alleviates tracking drift, and achieves stable tracking of badminton. The main differences between the models lie in their approaches to handling tracking problems and their adaptability to complex environments. The Faster RCNN-VGG19-OpenPose algorithm combines deep learning networks with human pose estimation technology, allowing it to more accurately capture the movement trajectory of a badminton shuttlecock, especially maintaining high tracking accuracy in situations where the shuttlecock is moving fast or there is occlusion. In

contrast, the VOT algorithm is primarily designed for tracking a single target and lacks the capability to track multiple targets, resulting in poor performance when dealing with multiple shuttlecocks or complex scenes. The MOT algorithm faces challenges in dealing with occlusions and similarities between targets, leading to a lower overlap between tracking and actual trajectories. Although the MCMOT algorithm has been improved for multi-camera environments, its performance is still not as good as the Faster RCNN-VGG19-OpenPose algorithm proposed in this study when dealing with fast-moving and complex backgrounds. To ensure the classification accuracy of the Faster RCNN-VGG19-OpenPose algorithm, the study also tests the classification accuracy of the Faster RCNN-VGG19-OpenPose algorithm for different badminton swing methods. The classification accuracy curve is shown in Fig. 10.

Fig. 10(a) presents the classification performance in the MPII. Fig. 10(b) shows the classification performance in the COCO dataset. From Fig. 10, the Faster RCNN-VGG19-OpenPose algorithm had the best classification performance in the two datasets. The best classification accuracy for badminton high and far balls, forehand strokes, forehand tennis balls, jump balls, and backhand picks reached 98.26%, 98.33%, 98.35%, 98.21%, and 97.08%, respectively, all exceeding 95%. Compared with the VOT algorithm, the Faster RCNN-VGG19-OpenPose algorithm improved the classification accuracy of high and far balls, forehand strokes, forehand tennis ball, jump balls, and backhand picks by about 18% to 25%. The above data indicates that the characteristics of the DT-SVM classifier in solving multi linear problems enable the Faster RCNN-VGG19-OpenPose algorithm to accurately classify badminton swing motions at each stage. Compared with other target tracking algorithms of the same type, it has more stable and superior recognition ability. The study also conducts multi-indicator tests on the above four algorithms using precision, recall, F1 value, and average detection time as indicators, as displayed in Table I.

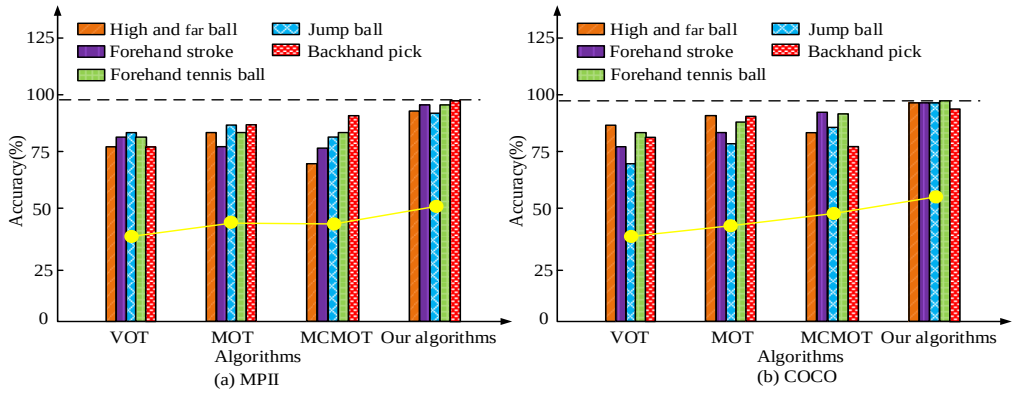


Fig. 10. Effectiveness of different algorithms in classifying badminton swing styles.

TABLE I. METRICS TEST RESULTS FOR VARIOUS ALGORITHMS

Data set	Algorithm	P/%	R/%	F1/%	Average detection time/s
MPII	VOT	88.86	87.38	88.57	5.18
	MOT	89.98	89.73	89.40	6.02
	MCMOT	91.87	88.64	90.26	5.38
	Faster RCNN-VGG19-OpenPose	95.41	94.28	95.85	2.74
COCO	VOT	89.28	88.16	88.37	4.61
	MOT	88.52	88.85	88.69	3.75
	MCMOT	91.43	90.79	91.11	3.08
	Faster RCNN-VGG19-OpenPose	95.56	93.68	95.32	2.51

According to Table I, among the four types of indicators detected, the VOT target tracking algorithm performed the worst, followed by the MOT target tracking algorithm, MCMOT target tracking algorithm, and the proposed Faster RCNN-VGG19-OpenPose algorithm. The highest P-value of the VOT target tracking algorithm was 89.28%, the highest R-value was 88.16%, the highest F1 was 88.37%, and the average detection time was 4.61s. The new Faster RCNN-VGG19-OpenPose target tracking algorithm proposed in the study had a maximum P-value of 95.56%, a maximum R-value of 93.68%, a maximum F1 value of 95.32%, and an average detection time of 2.51s. From this, the Faster RCNN-VGG19-OpenPose algorithm has relatively good performance, which is more suitable for the badminton tracking work at current stage.

B. Simulation Testing of Tracking and Motion Evaluation Model Based on Faster RCNN and VGG19

From the above test results, the Faster RCNN-VGG19-OpenPose algorithm performed excellently in badminton tracking and classification. However, this data is only feasible for the MPII and COCO datasets, and the persuasiveness and feasibility of the data results still need to

be further strengthened. The performance of the tracking and motion evaluation model based on Faster RCNN and VGG19 has not been verified yet. Therefore, the study attempts to use a self-made dataset for testing, which includes eight swing styles of badminton high and far balls, forehand strokes, forehand picks, forehand tennis ball, jump balls, backhand drives, backhand picks, and backhand forecourts. The number of videos for each badminton swing motion is 200, totaling 1400 video sequences. Each video sequence has a frame rate of 25fps, a resolution of 160x120, and an average length of 30s. At this point, the fuzzy comprehensive evaluation method is used to assign weights to the above motions. In addition, the study tracks the usage of GPU and CPU to record the time and storage space required for the model to process data. Tools such as Nsight are used to analyze resource utilization, adjust model parameters and structure to optimize efficiency, and test the long-term stability of the model in actual deployment to determine the resource consumption of each model. Table II displays the weighting results.

From Table II, the joint correlation under each motion, i.e. the weight score, was relatively reasonable and did not differ significantly from the actual physical sensation during motion. For example, in jump ball motions, there was a significant correlation between the weights of the legs, body, and knees. In forehand stroke, there is a significant correlation between the elbow, wrist, upper arm, and shoulder joints. The above motions are separately detected using confusion matrices. The same type popular motion recognition models are compared, including the LSTM model, Spatial Attention (SA), and Transformer network, as displayed in Fig. 11.

Fig. 11(a)-(d) show the confusion matrix results of LSTM, SA, Transformer, and the proposed model. From Fig. 11, the confusion matrix results for 7 different badminton swing motions under the LSTM model were poor, with only 4 groups scoring above 90 points. The same applies to the SA. There was a significant improvement in the confusion results of the Transformer, with 6 groups successfully paired for over 80 points. The confusion results of the model showed the best performance, with all 7 sets of badminton swing motions completed matching and scores above 90 points. Therefore, the designed method has certain effectiveness and performs better in similar models. In addition, to reflect the accuracy of the model in locating human skeletal points, the study also tests the model using the Percentage of Correct Key Points (PCK) and Frames Per Second (FPS) as indicators. The test results are shown in Fig. 12.

Fig. 12(a) shows the PCK performance of human bone point localization. Fig. 12(b) shows the FPS performance. As shown in Fig. 12, the proposed tracking and motion evaluation model based on Faster RCNN and VGG19 achieved an average FPS processing speed of 31.02 frames per second for five skeletal points in the human head, shoulder, elbow, wrist, and neck. Its highest PCK detection accuracy for the head, shoulder, elbow, wrist, and neck reached 98.05%, 98.10%, 97.89%, 97.55%, and 98.26%, respectively. The above data indicates that the new target tracking and motion evaluation model has unique advantages in recognizing human head, shoulder, elbow, wrist, and neck joint points. Finally, to explore the impact of the environment on the proposed model, the study also conducted comparative tests on the state-of-the-art motion assessment models under indoor and outdoor environments, using Mean Squared Error (MSE), Mean Absolute Error (MAE), resource consumption rate, and motion assessment accuracy as reference indicators. The models tested include detection methods based on the Deformable Parts Model (DFM), Tree-based Human Pose Estimation (TB-HPE), and Dual Source Deep Neural Network (DS-DNN) for human pose estimation. The test results are shown in Table III.

TABLE II. WEIGHTING VALUES FOR DIFFERENT MOTIONS AND JOINTS

Classification	High ball	Forehand stroke	Forehand pick	Forehand tennis ball	Jump ball	Backhand drive	Backhand pick	Backhand forecourt
Wrist	0.082	0.082	0.073	0.074	0.051	0.086	0.061	0.082
Elbow	0.083	0.074	0.085	0.072	0.053	0.087	0.078	0.079
Knee	0.035	0.042	0.059	0.052	0.067	0.044	0.047	0.048
Ankle	0.038	0.041	0.033	0.044	0.073	0.042	0.033	0.024
Hips and thighs	0.042	0.034	0.059	0.041	0.071	0.007	0.037	0.038
Crotch and body	0.043	0.043	0.048	0.036	0.082	0.024	0.024	0.035
Big arms and shoulders	0.074	0.067	0.054	0.051	0.011	0.055	0.038	0.031
Big arms and body	0.089	0.087	0.062	0.056	0.076	0.043	0.035	0.036

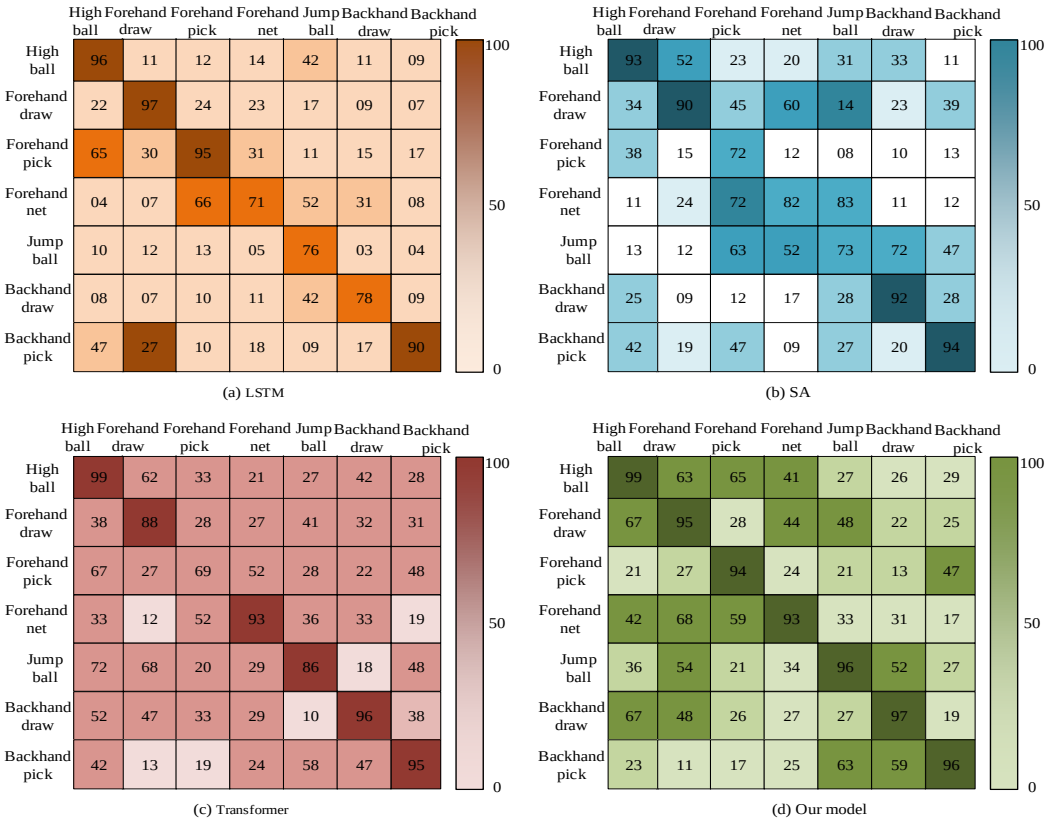


Fig. 11. Confusion matrix results for different recognition models.

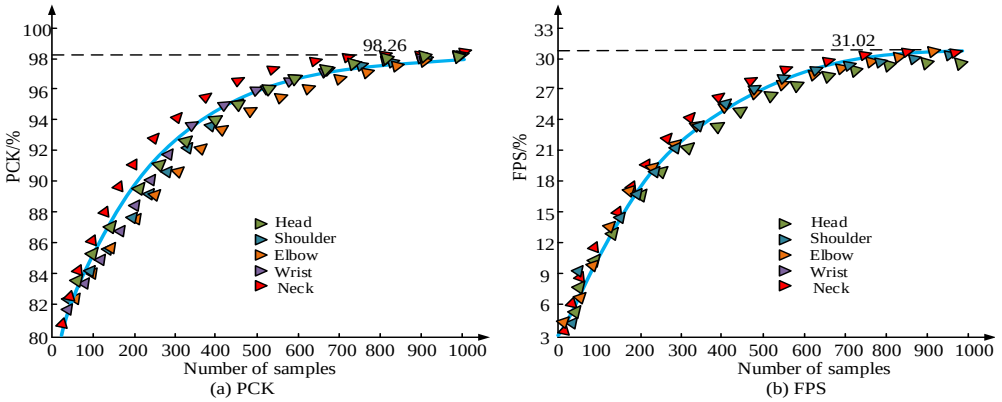


Fig. 12. The performance of modelling human skeletal point localization.

TABLE III. METRICS TEST RESULTS FOR DIFFERENT MODELS

Environments	Model	MSE	MAE	Resource consumption/%	Accuracy/s	References
Indoor	DFM	0.043	0.058	19.32	87.67	Lin L et al.
	TB-HPE	0.039	0.045	18.64	88.29	Gui W et al.
	DS-DNN	0.030	0.033	15.26	91.88	Xia Z et al.
	Our model	0.022	0.027	7.02	97.68	/
Outdoor	DFM	0.033	0.050	19.15	88.19	Lin L et al.
	TB-HPE	0.029	0.041	19.23	89.55	Gui W et al.
	DS-DNN	0.026	0.040	15.81	92.66	Xia Z et al.
	Our model	0.021	0.026	6.85	97.71	/

According to Table III, whether in indoor or outdoor environments, the DFM model had the worst performance in various indicators among the four models. The performance of TB-HPE, DS-DNN and the new target tracking and motion evaluation model increased from low to high. The proposed new target tracking and motion evaluation model had the lowest MSE value of 0.021, the lowest MAE value of 0.026, the lowest resource consumption rate of 6.85%, and the highest motion evaluation accuracy of 97.71%. Moreover, indoor and outdoor environments had almost no impact on the performance. In summary, the newly proposed model demonstrates the best overall performance and has more stable and superior recognition capabilities compared to the state-of-the-art motion assessment models currently available.

IV. Discussion

In the current field of sports motion target tracking and action assessment, deep learning technology, especially the application of Faster RCNN, has provided strong technical support for the automatic extraction of complex features from images. This enables algorithms to more accurately identify and track fast-moving sports targets, thereby significantly enhancing the accuracy and efficiency of target detection. However, due to the limited visual features of small-sized objects such as shuttlecocks, it is difficult for Faster RCNN to effectively distinguish them from similar small targets like sports shoes, light spots, and spectators' heads using limited visual features. In light of this, research has significantly improved Faster RCNN by introducing the VGG19 architecture, increasing the overall performance of the optimized Faster RCNN by about 11%. This indicates that the VGG19 architecture has a significant advantage in improving Faster RCNN's tracking and recognition of small targets. The result is consistent with the research findings of A Faghihi et al [13]. At the same time, in order to effectively obtain the posture information of the athletes, the study also used the improved Faster RCNN as the basic framework, and analyzed and processed the posture information of the badminton players through OpenPose and DT-SVM classifiers, finally proposing a tracking and action assessment model based on Faster RCNN and VGG19. Experimental results show that the proposed target tracking and action assessment model can achieve an average FPS processing speed of 31.02 frames per second for five key skeletal points of the human body: the head, shoulders, elbows, wrists, and neck. Moreover, its highest PCK detection accuracy rates for the head, shoulders, elbows, wrists, and neck reached 98.05%, 98.10%, 97.89%, 97.55%, and 98.26% respectively. This shows that OpenPose can process images in real-time and detect key points of multiple people, enhancing the model's joint point recognition ability. This is consistent with the research results of Chen C C et al. [31].

In summary, the research method has shown significant advantages in improving the accuracy, efficiency, and classification of badminton tracking and action assessment, which is consistent with the research conclusions of A Faghihi et al. and Chen C C et al., verifying the application potential and practical value of the method in target tracking and action assessment. Future work can further explore the integration of deep learning models, optimize algorithm efficiency, and expand the scope of applications to enhance the overall performance and applicability of sports motion analysis technology.

V. Conclusion

The research and development of sports tracking and motion evaluation have always been of great concern. The target recognition ability is crucial for athletes to achieve autonomous cooperation and optimized decision-making during the exercise process. In view of this, the Faster RCNN was used as the basic framework for target tracking. Then the VGG19 architecture and OpenPose algorithm were introduced for accuracy adjustment and pose estimation. Finally, a new badminton tracking and motion evaluation model based on Faster RCNN and improved VGG19 was proposed. Compared with other target tracking methods, the Faster RCNN-VGG19-OpenPose had the highest overlap between the real and tracking trajectories, which had good robustness, achieving stable tracking of badminton. The Faster RCNN-VGG19-OpenPose object tracking algorithm had a maximum P-value of 95.56%, R-value of 93.68%, and F1-value of 95.32%, with an average detection time of only 2.51s. It has relatively good performance, which is more suitable for the badminton tracking work at current stage. The simulation results showed that the proposed tracking and motion evaluation model based on Faster RCNN and VGG19 achieved an average FPS processing speed of 31.02 frames per second for five bone points in the human head, shoulder, elbow, wrist, and neck. Its highest PCK detection accuracy for the head, shoulder, elbow, wrist, and neck reached 98.05%, 98.10%, 97.89%, 97.55%, and 98.26%, respectively. The impact of the environment on the model was relatively small, and its overall performance was the best. Compared with similar motion recognition models, it has more stable and excellent recognition ability. In summary, the model has certain advantages and feasibility in the recognition and evaluation of badminton swing motions. However, this study only identifies a single badminton swing motion. Future research can add multiple combinations of badminton motions to improve the technical integrity. Additionally, although the study optimized the model through posture estimation technology, improving the estimation accuracy of the shoulder, elbow, and wrist joints, the estimation accuracy of other joint points has slightly decreased compared to the original model. To apply the research method to other more complex sports, future research can achieve more accurate human posture estimation by adjusting network structures, improving training strategies, or applying data augmentation techniques. At the same time, multimodal data, such as depth information or inertial sensor data, can be considered to enhance the model's ability to capture complex movements, thereby further improving the model's performance and providing more robust technical support for the rest of the sports training and assessment.

References

[1] Oh N. and Rodrigue H.. "Toward the development of large-scale inflatable robotic arms using hot air welding," Soft Rob, vol. 10, no. 1, pp. 88-96, January, 2023, DOI:10.1089/soro.2021.0134.

[2] RPohan M. A. and Utama J.. "Efficient Sampling-based for Mobile Robot Path Planning in a Dynamic Environment Based on the

Rapidly-exploring Random Tree and a Rule-template Sets," Int. J. Eng, vol. 36, no. 4, pp. 797-806, April, 2023, DOI:10.5829/IJE.2023.36.04A.16.

[3] Tian D., Han Y., Wang S., X. Chen and T. Guan. "Absolute size IoU loss for the bounding box regression of the object detection," NEUROCOMPUTING, vol. 500, no. 8, pp. 1029-1040, June, 2022, DOI:10.1016/j.neucom.2022.06.018.

[4] Meza J. , Delpiano J. , S. Velastín, R. Fernández and S. Awad. "Multiple Object Tracking for Robust Quantitative Analysis of Passenger Motion While Boarding and Alighting a Metropolitan Train," International Conference of Pattern Recognition Systems, vol. 43, no. 2, pp. 231-238, March, 2021, DOI:10.1049/icp.2021.1468.

[5] Shimizu T., Hachiuma R. and H. Kajita. "Hand Motion-Aware Surgical Tool Localization and Classification from an Egocentric Camera," Journal of Imaging,vol. 7, no. 2, pp. 15-19, March, 2021, DOI:10.3390/jimaging7020015.

[6] Li H. and Li D. "Recognition and Optimization Analysis of Urban Public Sports Facilities Based on Intelligent Image Processing," Hindawi, vol. 8, no. 9, pp. 42-48, March, 2021, DOI:10.1155/2021/8948248.

[7] Yin X, and Chen L.. "Image Object Detection Method Based on Improved Faster R-CNN," Journal of Circuits, Systems and Computers, vol. 33, no. 7, pp. 54-59, March, 2024, DOI:10.1142/S0218126624501305.

[8] Kandhro I. A., Manickam S. and Fatima K. "Performance evaluation of E-VGG19 model: Enhancing real-time skin cancer detection and classification," Heliyon, vol. 10, no. 10, pp. 35-41, March, 2024, DOI:10.1016/j.heliyon.2024.e31488.

[9] Awan M. J., Masood O. A. and Mohammed M. A.. "Image-Based Malware Classification Using VGG19 Network and Spatial Convolutional Attention," Electronics, vol. 10, no. 19, pp. 2444-2454, 2021, DOI:10.3390/electronics10192444.

[10] Verma S., Singh G. and Warishpat. "Detection of Traffic Sign using Inception V3 in Comparison with VGG-19 to Measure Accuracy," International Conference on Advance Computing and Innovative Technologies in Engineering, vol. 32, no. 4, pp. 730-733, March, 2023, DOI:10.1109/ICACITE57410.2023.10183243.

[11] Wan X., Zhang X. and L. Liu. "An Improved VGG19 Transfer Learning Strip Steel Surface Defect Recognition Deep Neural Network Based on Few Samples and Imbalanced Datasets," Applied Sciences, vol. 11, no. 6, pp. 2606-2616, March, 2021, DOI:10.3390/app11062606.

[12] Mohan R., Rama A. and K. Ganapathy. "Comparison of Convolutional Neural Network for Classifying Lung Diseases from Chest CT Images," International Journal of Pattern Recognition and Artificial Intelligence, vol. 8, no. 14, pp. 25-30, April, 2022, DOI:10.1142/S0218001422400031.

[13] Faghihi A., Fathollahi M. and R. Rajabi. "Diagnosis of skin cancer using VGG16 and VGG19 based transfer learning models," Multimedia Tools and Applications, vol. 83, no. 19, pp. 57495-57510, April, 2024, DOI:10.1007/s11042-023-17735-2.

[14] Ding L., Peng J. and Song L. "Automatically detecting apnea-hypopnea snoring signal based on VGG19+LSTM. Biomed," Signal Process. Control. vol. 80, no. 10, pp. 351-370, April, 2023, DOI:10.1016/j.bspc.2022.104351.

[15] Fan J., Yang X., Lu R., W. Li, and Y. Huang,"Long-term visual tracking algorithm for UAVs based on kernel correlation filtering and SURF features," Vis. Comput., vol. 39, no. 1, pp. 319-333, Jan. 2023. DOI: 10.1007/s00371-021-02331-y.

[16] Gali V., Babu B. C., R. B. Mutluri, M. Gupta, and S. K. Gupta,"Experimental investigation of Harris Hawk optimization-based maximum power point tracking algorithm for photovoltaic system under partial shading conditions," Opt. Control Appl. Methods, vol. 44, no. 2, pp. 577-600, Aug. 2023. DOI: 10.1002/oca.2773.

[17] He Q., Li X., and Li W. ,"Common Sports Injuries of Track and Field Athletes Using Cloud Computing and Internet of Things," Int. J. Comput. Intell. Syst., vol. 16, no. 1, p. 70, May 2023. DOI: 10.1007/s44196-023-00257-y.

[18] Murugan R. A., and Sathyabama B.,"Object Detection for Night Surveillance Using Ssan Dataset Based Modified Yolo Algorithm in Wireless Communication," Wireless Personal Commun., vol. 128, no. 3, pp. 1813-1826, Sep. 2023. DOI: 10.1007/s11277-022-10020-9.

[19] Chen Y., Zheng W., Zhao Y., Song T. H., and Shin H.,"Dw-yolo: an efficient object detector for drones and self-driving vehicles," Arabian J. Sci. Eng., vol. 48, no. 2, pp. 1427-1436, Feb. 2023a. DOI: 10.1007/s13369-022-06874-7.

[20] Chen Y., Xu H., X. Zhang, P. Gao, Z. Xu, and X. Huang,"An object detection method for bayberry trees based on an improved YOLO algorithm," Int. J. Digit. Earth, vol. 16, no. 1, pp. 781-805, 2023b Mar. DOI: 10.1080/17538947.2023.2173318.

[21] Yang Y., Chen L., Zhang J., Long L., and Wang Z., "UGC-YOLO: Underwater Environment Object Detection Based on YOLO with a Global Context Block," J. Ocean Univ. China, vol. 22, no. 3, pp. 665-674, May 2023. DOI: 10.1007/s11802-023-5296-z.

[22] Zhang J., He Y., Feng W., J. Wang, and N. N. Xiong,"Learning background-aware and spatial-temporal regularized correlation filters for visual tracking," Appl. Intell., vol. 53, no. 7, pp. 7697-7712, Jul. 2023. DOI: 10.1007/s10489-022-03868-8.

[23] Aygül K., Cikan M., Demirdelen T., and M. Tumay,"Butterfly optimization algorithm based maximum power point tracking of photovoltaic systems under partial shading condition," Energy Sources, Part A: Recovery, Util. Environ. Effects, vol. 45, no. 3, pp. 8337-8355, Oct. 2023. DOI: 10.1080/15567036.2019.1677818.

[24] Chessa A., Urso P. D', L. De Giovanni, V. Vitale, and A. Gebbia,"Complex networks for community detection of basketball players," Ann. Oper. Res., vol. 325, no. 1, pp. 363-389, Oct. 2023. DOI: 10.1007/s10479-022-04647-x.

[25] Rodríguez-Fernández A., R. Ramirez-Campillo, J. Raya-González, D. Castillo, and F. Y. Nakamura,"Is physical fitness related with in-game physical performance? A case study through local positioning system in professional basketball players," Proc. Inst. Mech. Eng., Part P: J. Sports Eng. Technol., vol. 237, no. 3, pp. 188-196, Jul. 2023. DOI: 10.1177/17543371211031160.

[26] Rahimian P. and Toka L.,"Optical tracking in team sports: A survey on player and ball tracking methods in soccer and other team sports," J. Quant. Anal. Sports, vol. 18, no. 1, pp. 35-57, Mar. 2022. DOI: 10.1515/jqas-2020-0088.

[27] Liu H., Duan X., H. Chen, H. Lou, and L. Deng,"DBF-YOLO: UAV Small Targets Detection Based on Shallow Feature Fusion," IEEJ Trans. Electr. Electron. Eng., vol. 18, no. 4, pp. 605-612, Jan. 2023. DOI: 10.1002/tee.23758.

[28] Gai R., Chen N., and H. Yuan,"A detection algorithm for cherry fruits based on the improved YOLO-v4 model," Neural Comput. Appl., vol. 35, no. 19, pp. 13895-13906, Jul. 2023. DOI: 10.1007/s00521-021-06029-z.

[29] Yang D., "Research on multi-target tracking technology based on machine vision," Appl. Nanosci., vol. 13, no. 4, pp. 2945-2955, Apr. 2023. DOI: 10.1007/s13204-021-02293-6.

[30] Hasanvand M., Nooshyar M., E. Moharamkhani, and A. Selyari. "Machine Learning Methodology for Identifying Vehicles Using Image Processing," AIA, vol. 1, no. 3, pp. 170-178, Apr, 2023, DOI: https://doi.org/10.47852/bonviewAIA3202833

[31] Chen C. C., Chang C., Lin C. S., et al., "Video Based Basketball Shooting Prediction and Pose Suggestion System," Multimedia Tools and Applications, vol. 82, no. 18, pp. 27551-27570, May 2023. Doi: 10.1007/s11042-023-14490-2.

Cyberbullying Detection on Social Networks Using a Hybrid Deep Learning Architecture Based on Convolutional and Recurrent Models

Aigerim Altayeva[1], Rustam Abdrakhmanov[2], Aigerim Toktarova[3], Abdimukhan Tolep[4]
International Information Technology University, Almaty, Kazakhstan[1]
International University of Tourism and Hospitality, Turkistan, Kazakhstan[2]
M. Auezov South Kazakhstan University, Shymkent, Kazakhstan[3]
Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkistan, Kazakhstan[4]

***Abstract*—This research paper explores the development and efficacy of a hybrid deep learning architecture for cyberbullying detection on social media platforms, integrating Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. By leveraging the strengths of both CNNs and LSTMs, the model aims to enhance the accuracy and sensitivity of detecting cyberbullying incidents. The study systematically evaluates the performance of the proposed model through a series of experiments involving a diverse dataset derived from various social media interactions, categorized by sentiment and type of bullying. Results indicate that while the model achieves high accuracy in identifying cyberbullying, challenges such as overfitting and the need for better generalization to unseen data persist. The paper also discusses ethical considerations and the potential for bias in automated monitoring systems, stressing the importance of ethical AI practices in social media governance. The findings underscore the complexity of automated cyberbullying detection and highlight the necessity for advanced machine learning techniques that are robust, scalable, and aligned with ethical standards. This study contributes to the broader discourse on the application of artificial intelligence in enhancing digital safety and advocates for a multidisciplinary approach to address the socio-technical challenges posed by cyberbullying in the digital age.**

Keywords*—*Cyberbullying detection; deep learning; CNN; LSTM; social media monitoring; sentiment analysis; digital safety

I. Introduction

Cyberbullying has emerged as a significant social problem with the rise of digital communication platforms. Unlike traditional forms of bullying, cyberbullying allows perpetrators to extend their reach beyond physical spaces, affecting victims at any time and from any location. This form of harassment is especially prevalent among adolescents and young adults and has been linked to a range of negative outcomes, including depression, anxiety, and even suicidal thoughts [1]. The anonymous and pervasive nature of online interactions complicates the detection and intervention processes, making it imperative to develop automated tools that can effectively identify and mitigate instances of cyberbullying.

Recent advancements in machine learning and natural language processing have paved the way for more sophisticated approaches to monitoring online behavior. In particular, deep learning architectures have demonstrated significant potential in text classification tasks, which are central to identifying harmful content [2]. However, the complexity of language, including the use of slang, coded messages, and context-specific references, poses unique challenges that require robust and adaptable models.

Hybrid deep learning architectures that combine convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have shown promising results in various domains, such as image recognition and speech processing [3]. CNNs are adept at extracting hierarchical feature maps from spatial data, making them useful for capturing the textural patterns in language, whereas RNNs, particularly those using Long Short-Term Memory (LSTM) cells, excel in handling sequences with long-range dependencies, capturing the contextual nuances necessary for understanding the intent behind words [4].

The proposed research aims to leverage the strengths of both CNNs and LSTMs to develop a comprehensive model for cyberbullying detection on social networks. This approach is motivated by the hypothesis that a hybrid model can better accommodate the complex and dynamic nature of language used in online social platforms than models based solely on one type of network. By integrating the local feature extraction capabilities of CNNs with the sequence modeling strengths of LSTMs, the model is expected to perform with higher accuracy and sensitivity in detecting cyberbullying instances [5].

Additionally, the vast amount of data available on social networks provides a fertile ground for training deep learning models but also introduces challenges such as data imbalance, privacy concerns, and the need for models to perform well across diverse demographic groups. Addressing these challenges requires not only technical solutions but also an understanding of the social and ethical implications of deploying AI-driven monitoring tools [6].

The use of dropout layers and other regularization techniques in the architecture aims to prevent overfitting, ensuring that the model generalizes well to new, unseen data, which is critical in the ever-evolving landscape of social media language [7]. Moreover, the model's performance must be rigorously evaluated to ensure its efficacy and fairness,

preventing potential biases that could harm certain groups disproportionately [8].

This research contributes to the growing body of knowledge on AI applications in social good by providing a nuanced approach to detecting cyberbullying, which could be integrated into social networks as part of proactive measures to safeguard users. The outcome of this study has the potential to inform policies and practices not only for social media companies but also for educators and policymakers concerned with cyber welfare [9].

In summary, the increasing prevalence of cyberbullying and the limitations of current detection methodologies underscore the need for innovative solutions. This paper proposes a novel hybrid deep learning architecture that integrates the distinct advantages of CNNs and LSTMs, aiming to enhance the detection accuracy and operational efficiency of cyberbullying interventions on social platforms. By addressing both technical challenges and ethical considerations, this study seeks to contribute significantly to the safer utilization of digital spaces.

II. Related Work

The study of cyberbullying has garnered significant interest due to its profound impact on individuals and society. This section delves into the evolution of cyberbullying detection methodologies, tracing the trajectory from manual monitoring to sophisticated hybrid deep learning models. By examining past efforts and their limitations, this paper aims to underscore the novelty and necessity of the proposed approach in enhancing detection mechanisms on social media platforms [10].

A. Cyberbullying: Definitions and Impact

Cyberbullying is characterized by repeated behavior aimed at harming others using digital platforms. Definitions vary, but common elements include intent, repetition, and the use of electronic forms of contact [11]. As for its impact, studies have consistently shown that victims of cyberbullying exhibit higher levels of anxiety, depression, and even suicidal ideation, demonstrating the critical need for effective detection and prevention strategies [12].

B. Traditional Methods for Cyberbullying Detection

Initially, cyberbullying detection relied heavily on manual monitoring, where human moderators reviewed content to identify harmful behavior. However, this method is not scalable and is subject to human error and bias [13]. Subsequently, rule-based systems were developed, utilizing predefined keywords and patterns to automate detection. These systems, while faster, often failed to capture the context and complexity of human interactions, leading to high rates of false positives and negatives [14].

C. Machine Learning Approaches

The advent of machine learning heralded a new era in cyberbullying detection, with the development of sophisticated, feature-based models. Support Vector Machines (SVM) and Naive Bayes classifiers are prominent examples, employed extensively to analyze textual data. These models classify text by extracting and leveraging specific features such as word frequency and sentence structure [15]. Despite their initial success, these traditional machine learning approaches require extensive feature engineering to operate effectively. Moreover, their rigid frameworks often fail to adapt to the dynamic and evolving use of language typical of social media contexts. This limitation significantly restricts their practical applicability, as they cannot seamlessly accommodate new slang, code-switching, or emergent linguistic patterns without manual updates or retraining [16].

D. Deep Learning in Text Analysis

The limitations of traditional machine learning models led to the adoption of deep learning techniques, which can automatically learn features from data. Single-model architectures, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have been widely used in text analysis due to their ability to extract spatial hierarchies and manage sequence dependencies, respectively [17]. CNNs are effective in extracting text patterns without the need for manual feature specification, while RNNs, especially those equipped with LSTM cells, excel in understanding context over longer text sequences [18]. However, each has limitations when used independently; CNNs can miss nuanced linguistic context, and RNNs can be computationally intensive and difficult to train on long sequences [19].

E. Hybrid Deep Learning Models

The concept of hybrid deep learning models, which integrate the strengths of CNNs and LSTMs, has started to gain traction. These models capitalize on CNNs' ability to process spatial data and LSTMs' proficiency in sequence processing, offering a robust framework for understanding complex text data. Such architectures have shown promise in fields like sentiment analysis and natural language processing, where both local features and global context are critical for accurate interpretation [20]. In the realm of cyberbullying detection, these hybrid models are proposed to enhance accuracy by effectively capturing both the textual nuances and the broader semantic context of interactions [21].

F. Challenges and Ethical Considerations

Implementing deep learning models for cyberbullying detection on social networks introduces several challenges. Data imbalance—where instances of bullying are vastly outnumbered by normal interactions—can skew model training and performance [22]. Privacy is another significant concern, as these models require access to personal data, potentially infringing on user confidentiality [23]. Moreover, the deployment of such models must be handled with care to avoid biases that could disproportionately affect certain demographics, necessitating continuous evaluation to ensure fairness and ethical integrity [24].

G. Research Gaps and Opportunities

Despite the advancements in detection technologies, several research gaps remain. Current models often fail to consider the multi-modal nature of cyberbullying, which can include text, images, and videos. Moreover, the adaptability of models to new, evolving forms of language and bullying tactics is still limited. There is a pressing need for models that can dynamically learn and adapt from incoming data streams without requiring frequent retraining [25]. Furthermore, there is a lack of comprehensive studies that integrate ethical

considerations into the model development process, an area ripe for further exploration [26-28].

III. Materials and Methods

A. Proposed Model

The proposed hybrid deep learning architecture integrates Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to exploit the strengths of both in detecting cyberbullying within textual data. This section elaborates on the configuration, interplay, and computational aspects of the model. Fig. 1 demonstrates the proposed model for cyberbullying detection.

The model comprises several layers, each tailored to perform specific functions critical to processing and classifying textual inputs:

Input Layer: This layer accepts pre-processed textual data, where each input unit represents a tokenized word encoded as a vector in a high-dimensional space.

Dropout Layer: Positioned immediately after the input layer, this layer randomly omits a fraction p of the input units (neurons) during training to prevent overfitting, where p is the dropout rate set to 0.5 as a starting point.

LSTM Layer: The LSTM layer processes the sequence data, preserving long-term dependencies within texts. The LSTM cells are formulated as follows:

$$i_t = \sigma\left(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}\right) \quad (1)$$

$$f_t = \sigma\left(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf}\right) \quad (2)$$

$$g_t = \tanh\left(W_{ig}x_t + b_{ig} + W_{hg}h_{t-1} + b_{hg}\right) \quad (3)$$

$$o_t = \sigma\left(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho}\right) \quad (4)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes g_t \quad (5)$$

$$h_t = o_t \otimes \tan g\left(c_t\right) \quad (6)$$

Where, I, f, g, and o are the input, forget, cell, and output gates, respectively; W and b represent weights and biases; σ denotes the sigmoid function; and ⊙ indicates element-wise multiplication.

Convolutional Layer: Following the LSTM, a convolutional layer is applied to extract local feature patterns from the output sequences of the LSTM, enhancing the model's ability to detect nuanced language patterns indicative of cyberbullying. The convolutional operation is defined as:

$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a) \quad (7)$$

where x is the input from the LSTM, w is the kernel, and s(t) is the convolved feature at time t.

Flatten Layer: This layer flattens the multi-dimensional output of the convolutional layer into a single-dimensional array for subsequent processing.

Output Layer: The final layer is a fully connected layer that maps the flattened features to the output classes, which are 'cyberbullying' and 'non-cyberbullying'. The activation function used here is the softmax function, given by:

$$Soft\max(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (8)$$

where z_i are the inputs to the output layer from the flatten layer.

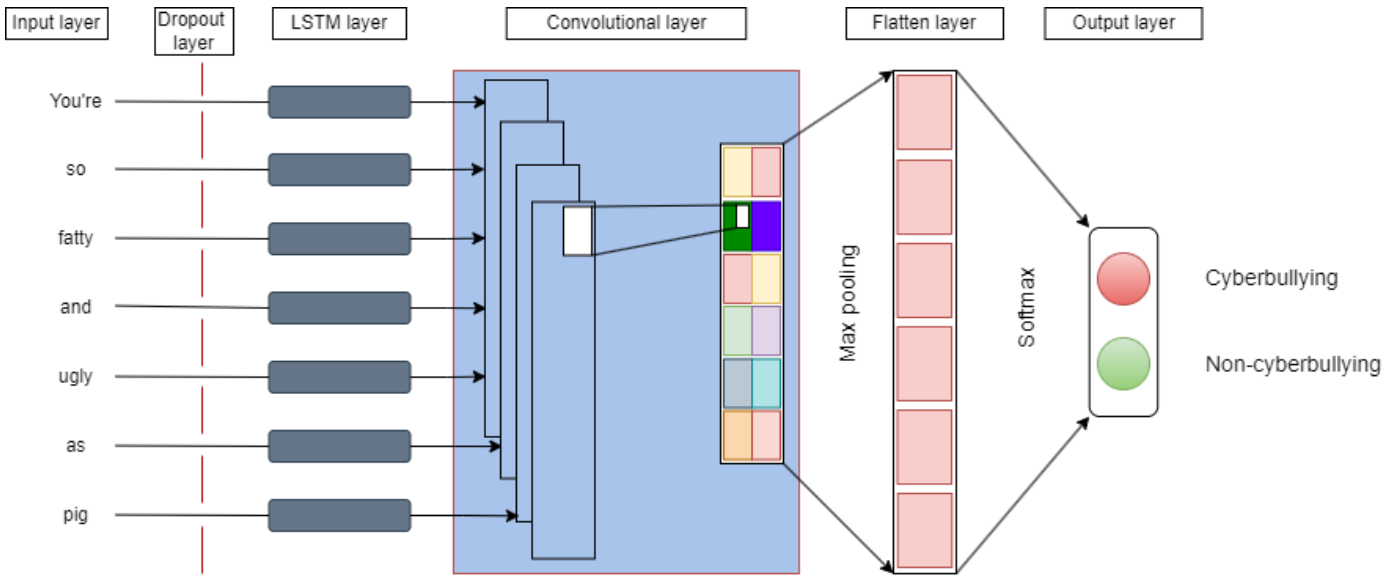


Fig. 1. Architecture of the proposed model.

The proposed model is trained using a cross-entropy loss function, which is suitable for binary classification problems like cyberbullying detection. The loss is calculated as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right] \quad (9)$$

Where, N is the number of training samples, y_i is the actual label, and $\hat{y}_i$ is the predicted probability of the sample belonging to the 'cyberbullying' class.

By combining LSTM and CNN layers, the architecture harnesses both temporal dynamics and spatial feature extraction, making it robust against the complex variations in textual data typical of cyberbullying contexts. This model promises not only improved accuracy but also better generalizability across different social media platforms.

B. Dataset

The dataset utilized in this study is sourced from Kaggle and consists of structured data specifically tailored for cyberbullying detection tasks. This dataset is comprised of textual data collected from various social media platforms, providing a diverse representation of linguistic styles and contexts. It includes examples labeled with categories indicative of different forms of cyberbullying, facilitating the training and evaluation of machine learning models designed to identify and classify cyberbullying instances.

The data are categorized into distinct classes based on the nature and severity of the bullying behavior. This classification aids in fine-tuning the model's sensitivity to various forms of cyberbullying, from subtle to overt aggression. The dataset's structured format and comprehensive labeling make it an ideal resource for developing robust detection systems using advanced machine learning techniques such as the proposed hybrid deep learning architecture.

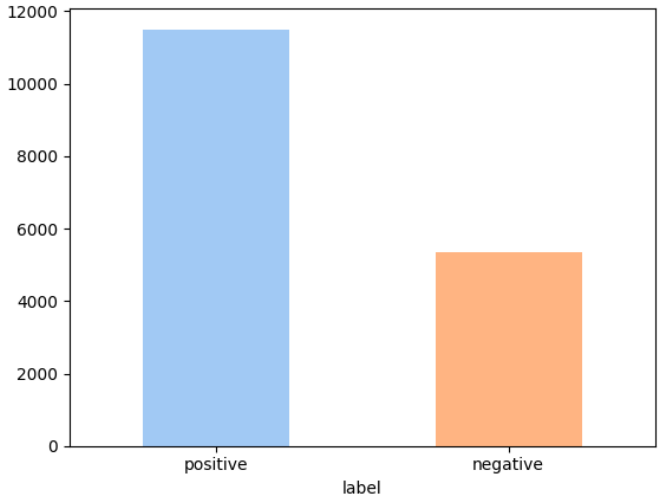


Fig. 2. Class distribution in the cyberbullying dataset.

This dataset not only enhances the model's ability to learn from real-world examples but also provides a benchmark for assessing the effectiveness of different architectural configurations and learning algorithms in accurately detecting cyberbullying across diverse digital environments.

In Fig. 2, the dataset depicted in the provided bar chart is characterized by a significant imbalance between two classes: "positive" and "negative". The "positive" class, which presumably represents instances identified as cyberbullying, substantially outnumbers the "negative" class, indicative of non-cyberbullying instances. This distribution highlights a common challenge in machine learning applications, where class imbalance can significantly impact the learning algorithm's ability to accurately generalize to new data. Such disparities necessitate the implementation of specialized techniques in data preprocessing, such as oversampling the minority class or undersampling the majority class, to ensure that the model trained on this dataset does not exhibit a bias towards the more frequently represented class. This imbalance also underscores the importance of employing robust evaluation metrics that can appropriately measure model performance in the presence of skewed class distributions.

IV. Results

A. Evaluation Parameters

In the context of this research paper on cyberbullying detection using a hybrid deep learning architecture, the evaluation of the model's performance is critical. This involves several key parameters that offer insights into the effectiveness and reliability of the model across various dimensions:

Accuracy measures the proportion of total correct predictions (both positive and negative) over all predictions made by the model. This metric is straightforward but can be misleading in datasets where class imbalance exists, as it might reflect the prevalence of the majority class rather than the model's ability to identify all classes accurately.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

Precision assesses the correctness of positive predictions. In the context of cyberbullying, it measures the proportion of true cyberbullying instances among all instances flagged by the model as cyberbullying. High precision is crucial in this domain to minimize the number of non-bullying content mistakenly classified as bullying, which can have significant social and psychological impacts on users.

$$preision = \frac{TP}{TP + FP} \quad (11)$$

Recall, also known as sensitivity, recall quantifies the model's ability to identify all actual positives. It answers the question: "Of all the true cyberbullying instances, how many did the model correctly identify?" High recall is essential for ensuring that the model does not overlook cyberbullying instances, thus providing a safe online environment.

$$recall = \frac{TP}{TP + FN} \quad (12)$$

The F-score or F1-score is the harmonic mean of precision and recall. It is particularly useful when the balance between precision and recall is important to the application. The F-score provides a single metric that summarizes model performance in terms of both the precision and the recall, which is valuable in scenarios where both false positives and false negatives have serious implications.

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (13)$$

These metrics provide a comprehensive view of the model's performance, allowing for balanced optimization in scenarios where making accurate and reliable predictions is crucial. In cyberbullying detection, where the cost of false positives (non-bullying content flagged as bullying) and false negatives (failing to detect actual bullying) can be high, these measures help in fine-tuning the model to achieve optimal performance.

B. Data Preparation Results

Fig. 3 illustrates a word cloud generated from a topic modeling analysis on the cyberbullying dataset, depicting the most frequently occurring words within a specific cluster of the data labeled as "Topic 1." The prominent terms, including derogatory and offensive language, highlight the aggressive and harmful nature of the communications that are commonly identified in instances of cyberbullying. Words such as "dumb," "racist," and more severe pejorative terms point to the type of content that algorithms need to effectively identify and address. This visualization not only underscores the challenges of automatically detecting such harmful content but also serves as a stark reminder of the harsh reality faced by victims of cyberbullying. The word cloud thus provides a visual summary of the key themes and language used in cyberbullying, guiding the development of more nuanced detection mechanisms.

Fig. 3. Word cloud from cyberbullying dataset.

Fig. 4 presents the sentiment distribution across various topics derived from the cyberbullying dataset, segmented into negative, neutral, and positive sentiments. The vertical stacked bar graph illustrates a notable predominance of negative sentiments across all topics, underscoring the pervasive nature of negative expressions in discussions related to cyberbullying. Topic 0 exhibits a particularly high concentration of negative sentiment, dwarfing the other sentiments and suggesting a significant alignment with more aggressive or harmful content. In contrast, Topics 1 through 3 display a more balanced distribution but still with a visible skew towards negative sentiments. This visualization highlights the critical need for effective sentiment analysis tools in cyberbullying detection systems, as it underscores the varying emotional contexts that can pervade different discussion topics, aiding in more nuanced understanding and classification of potentially harmful content.

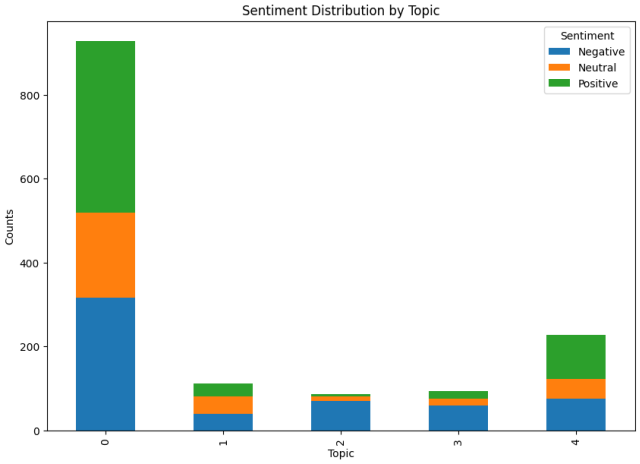


Fig. 4. Distribution of classes in the dataset.

Fig. 5 depicts the sentiment distribution across five different topics extracted from a cyberbullying dataset, with sentiments categorized as negative, neutral, and positive. The heat map highlights the count of sentiments within each topic, revealing a complex landscape of emotional expressions. Notably, Topic 0 is predominantly characterized by positive sentiments, contrasting sharply with other topics where negative sentiments prevail. This suggests that Topic 0 might encompass discussions or interactions of a different nature or context compared to the others. Topics 1 through 4 show a higher prevalence of negative sentiments, indicative of the adversarial or harmful content typically associated with cyberbullying. This distribution aids in understanding the emotional underpinnings of the topics discussed within the dataset and underscores the importance of sentiment analysis in contextualizing the interactions captured in social media data for cyberbullying detection.

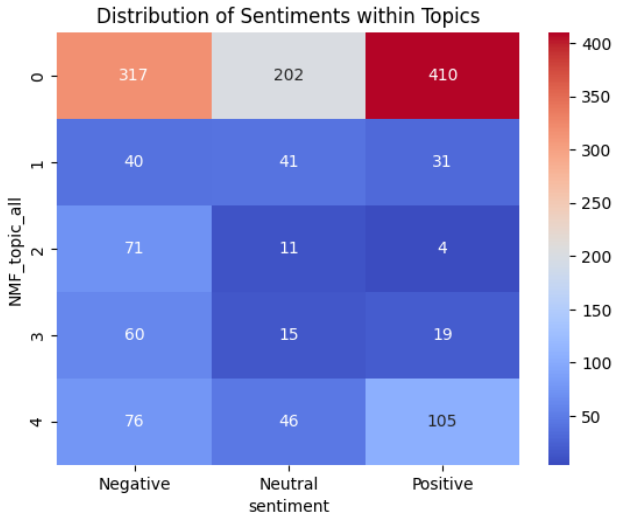


Fig. 5. Distribution of sentiments within topics.

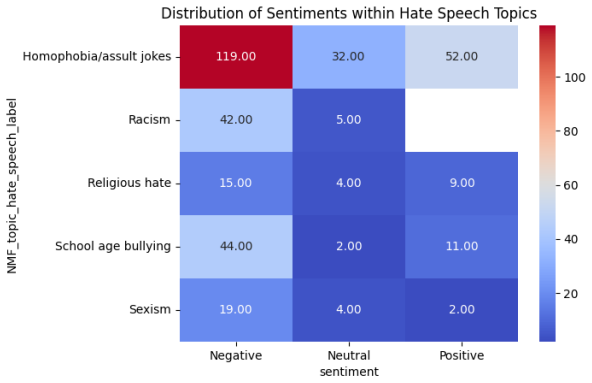


Fig. 6. Distribution of sentiments within hate speech topics.

Fig. 6 presents a heat map illustrating the distribution of sentiments categorized as negative, neutral, and positive across various hate speech topics derived from social media data. This visualization reveals that homophobia and assault jokes have a notably high incidence of positive sentiments relative to their negative sentiments, suggesting complex interactions that might include endorsements or trivialization of serious issues. In contrast, topics like racism, religious hate, school-age bullying, and sexism predominantly exhibit negative sentiments, aligning with the adversarial nature of the content. The presence of neutral sentiments across all topics indicates discussions that potentially include factual statements or undetermined stances. This detailed sentiment breakdown provides valuable insights into the nature of discussions within each hate speech category, underscoring the importance of nuanced analysis in automated detection systems to differentiate between contexts and intentions behind the words used.

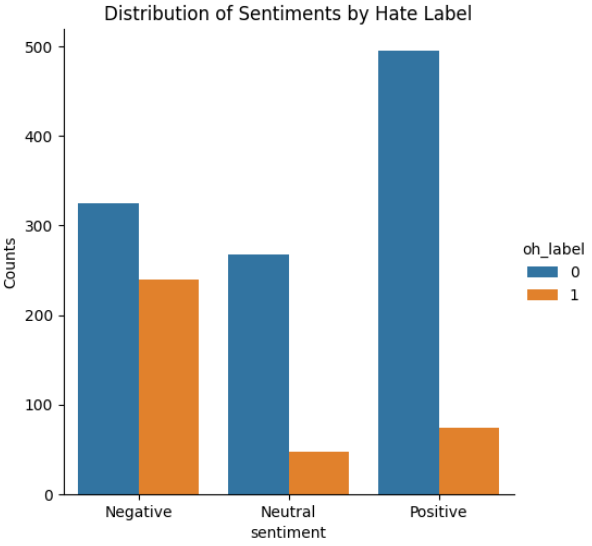


Fig. 7. Distribution of sentiments by hate label.

Fig. 7 displays a bar chart representing the distribution of sentiments categorized as negative, neutral, and positive across two distinct hate labels, identified here as "0" and "1". The labels likely denote the presence (1) or absence (0) of hate speech within the sentiments expressed. Notably, there is a predominant concentration of negative sentiments associated with the label "1", which underscores the alignment of negative sentiments with recognized hate speech. In contrast, sentiments labeled as "0", suggesting non-hate speech contexts, show a more balanced distribution among negative, neutral, and positive sentiments. This distribution highlights the complex relationship between expressed sentiments and their classification as hate speech, illustrating the challenge in distinguishing hate speech from non-hate speech based on sentiment alone. The data underscores the importance of sophisticated models that can interpret not only the sentiment but the context and intent behind expressions to effectively moderate content on social platforms.

The analysis presented in this subsection highlights the intricate relationship between sentiment and hate speech within digital communications. The visualizations and data distributions discussed reveal not only the prevalence of negative sentiments associated with confirmed instances of hate speech but also the more nuanced interplay of sentiments in texts labeled as non-hate speech. These insights underscore the challenges in deploying automated systems for the detection and moderation of hate speech, emphasizing the necessity for models that can adeptly discern context, intent, and sentiment nuances. Future research should focus on enhancing the accuracy and sensitivity of detection algorithms by incorporating multidimensional data analysis and applying advanced machine learning techniques that can better understand and process the complexities of human language and interaction in digital platforms. This approach is crucial for developing more effective and ethical tools for managing online safety and promoting positive digital communication environments.

C. Experimental Results

Fig. 8 presents a confusion matrix for a sentiment classification model applied to cyberbullying detection, displaying the numbers of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions. In the matrix, the horizontal axis represents the actual labels of the data ('Negative' and 'Positive'), while the vertical axis represents the predicted labels. The matrix shows that the model correctly identified 1498 negative cases and 531 positive cases as true negatives and true positives, respectively. Conversely, there were 228 cases where positive sentiments were incorrectly predicted as negative (false negatives), and 271 cases where negative sentiments were incorrectly labeled as positive.

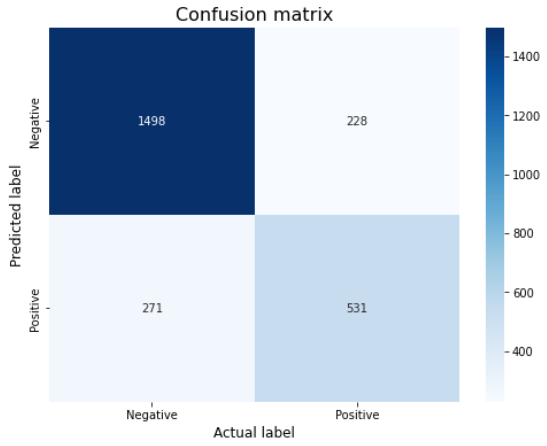


Fig. 8. Confusion matrix of sentiment classification model.

The significant number of true positives and true negatives indicates that the model is fairly effective in identifying both negative and positive sentiments accurately. However, the presence of false positives and false negatives highlights areas for potential improvement. The false negatives, in particular, are concerning because they represent instances where the model failed to detect positive sentiments, which could be crucial in contexts where affirming positive communication is as important as detecting negative or harmful content. Enhancing the model's sensitivity and specificity, possibly through more advanced feature engineering, optimization of hyperparameters, or the incorporation of context-aware algorithms, could further improve its performance. This analysis not only serves as a basis for evaluating the effectiveness of the current model but also guides future modifications and enhancements to better handle the complexities of sentiment analysis in cyberbullying contexts.

Fig. 9 displays the training and validation accuracy and loss of a deep learning model over several epochs. The left graph illustrates a steady increase in training accuracy from approximately 82.5% to 95.4% across six epochs, suggesting that the model is effectively learning from the training data. However, the validation accuracy starts around 90% and shows a declining trend to stabilize near 87.5%, indicating potential issues with overfitting as the model becomes too specialized to the training data and performs less effectively on unseen validation data. This divergence between training and validation performance is a common challenge in machine learning, highlighting the need for techniques like regularization or dropout to mitigate overfitting.

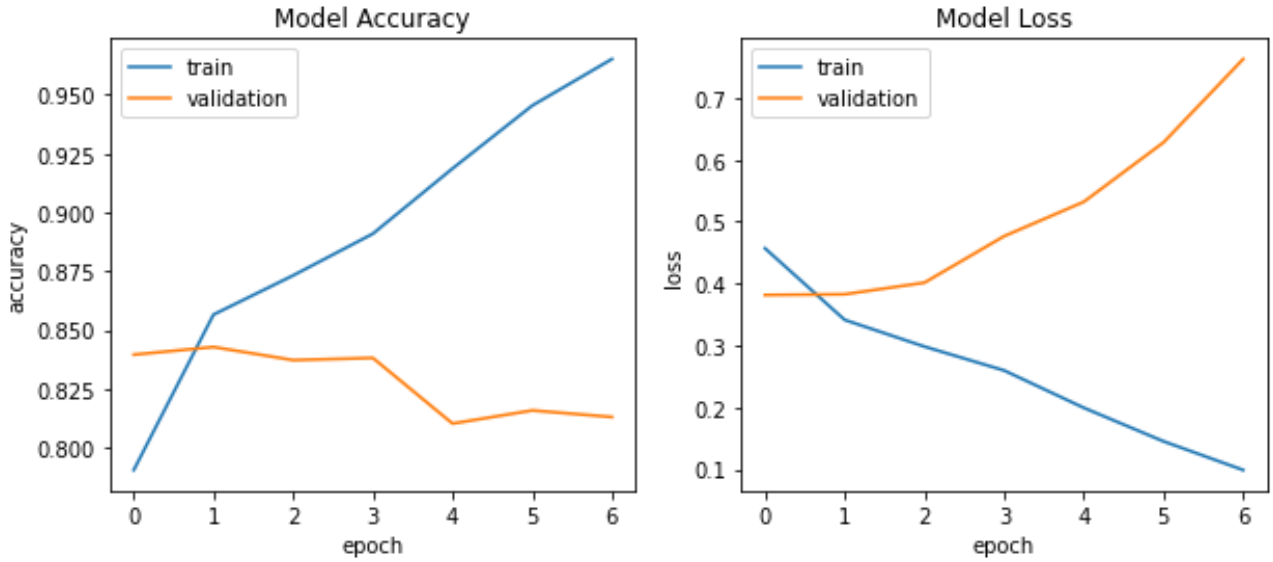


Fig. 9. Model accuracy and loss.

The right graph of model loss shows an inverse relationship to the accuracy graph. The training loss decreases sharply from about 0.7 to under 0.1, while the validation loss initially decreases but then begins to rise after the first epoch, reaching about 0.6 by the sixth epoch. This increase in validation loss alongside a decrease in validation accuracy further supports the possibility of overfitting. The early plateau and subsequent increase in validation loss could also suggest that the model's capacity is either insufficient or excessively tuned to nuances in the training data that do not generalize well. Addressing this might involve adjusting model parameters, introducing early stopping in training, or enriching the dataset to enhance the model's generalization capabilities. These adjustments are crucial for improving model robustness and ensuring reliable predictions across diverse data sets.

V. Discussion

The development and evaluation of a hybrid deep learning model to detect cyberbullying across social media platforms have yielded significant insights and highlighted several areas of concern and opportunity that warrant further discussion. The integration of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks in the proposed model aimed to leverage the spatial and temporal processing strengths of these architectures to enhance detection accuracy.

A. Model Performance and Overfitting

The proposed system demonstrated significant accuracy in the training and validation results, as illustrated in Fig. 9, demonstrate a clear disparity between training accuracy and validation accuracy, suggesting a tendency towards overfitting. This phenomenon, where the model performs exceptionally well on training data but less so on unseen validation data, is indicative of the model's inability to generalize beyond the training set [29-31]. Overfitting is a common challenge in machine learning, particularly in complex models such as deep neural networks. Strategies such as introducing dropout layers, increasing data augmentation, and employing regularization techniques have been recommended to mitigate this issue [32-34].

B. Sentiment and Hate Speech Analysis

The analysis of sentiments and their correlation with hate speech, as discussed through Fig. 5, 6, and 7, reveals a complex interaction between sentiment classifications and the

identification of hate speech. The prevalence of negative sentiments in recognized hate speech categories underscores the critical role of sentiment analysis in contextualizing content. However, the presence of positive and neutral sentiments within these categories also highlights the complexity of human communication, where not all harmful or aggressive communications are overtly negative [35-37].

C. Ethical Considerations and Bias

The deployment of AI in monitoring and moderating online content raises substantial ethical questions, particularly concerning privacy, consent, and the potential for bias. The model's susceptibility to biases, whether in data representation or algorithmic predilections, can lead to disproportionate flagging of content from specific demographics or linguistic groups [38-40]. It is crucial that continuous efforts be made to audit and refine these systems to ensure fairness and equity in automated content moderation [41].

D. Future Directions

Looking forward, the research highlights several pathways for improvement and further study. First, enhancing the model's generalizability through a more diverse and extensive dataset could help in reducing overfitting and improving the model's robustness across various social media contexts [42]. Second, exploring alternative hybrid architectures or newer approaches like transformer models, which have shown promise in other NLP tasks, might offer new avenues for increasing both accuracy and efficiency in processing [43-45].

Additionally, the integration of multimodal data, including images and videos alongside text, could provide a more holistic view of content, given that cyberbullying often transcends textual communication [46-47]. Finally, more nuanced understanding and classification frameworks are necessary to better differentiate between types of cyberbullying and their severities, which could aid in more targeted and effective interventions.

This research contributes to the ongoing discourse on AI's role in social media safety and the technical challenges associated with developing effective cyberbullying detection systems. While the results are promising, they also reflect the need for a cautious and considered approach to implementing such technologies. The balance between technological advancements and ethical considerations remains a critical frontier in the research and application of AI in societal contexts.

VI. Conclusion

The research undertaken in this paper has highlighted the pivotal role that hybrid deep learning architectures can play in the detection and analysis of cyberbullying across social media platforms. The integration of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks demonstrated potential in enhancing the accuracy of identifying harmful content by leveraging both spatial and sequential data processing capabilities. Despite this, challenges such as model overfitting and the need for greater generalization point to areas requiring further refinement. Ethical considerations around the implementation of AI in public domains, particularly concerning privacy, bias, and fairness, underscore the necessity for rigorous standards and continuous oversight in deploying such technologies. Future research directions should focus not only on improving the technical robustness of detection models through advanced machine learning techniques and diversified data sets but also on exploring the socio-technical impacts of AI interventions in human interactions. The balance between technological innovation and ethical responsibility remains a critical frontier in the application of AI, advocating for a multidisciplinary approach to ensure that advancements in AI contribute positively to societal welfare. This research contributes to a growing body of knowledge that seeks to harness the power of AI for social good, emphasizing the importance of aligning technological developments with human values and ethical standards.

Acknowledgment

This work was supported by the Science Committee of the Ministry of Higher Education and Science of the Republic of Kazakhstan within the framework of grant AP23488900 "Automatic detection of cyberbullying among young people in social networks using artificial intelligence".

References

[1] Murshed, B. A. H., Abawajy, J., Mallappa, S., Saif, M. A. N., & Al-Ariki, H. D. E. (2022). DEA-RNN: A hybrid deep learning approach for cyberbullying detection in Twitter social media platform. IEEE Access, 10, 25857-25871.

[2] Sultan, D., Mendes, M., Kassenkhan, A., & Akylbekov, O. (2023). Hybrid CNN-LSTM Network for Cyberbullying Detection on Social Networks using Textual Contents. International Journal of Advanced Computer Science and Applications, 14(9).

[3] Albayari, R., Abdallah, S., & Shaalan, K. (2024). Cyberbullying Detection Model for Arabic Text Using Deep Learning. Journal of Information & Knowledge Management, 2450016.

[4] Omarov, B., Narynov, S., & Zhumanov, Z. (2023). Artificial Intelligence-Enabled Chatbots in Mental Health: A Systematic Review. Computers, Materials & Continua, 74(3).

[5] Iwendi, C., Srivastava, G., Khan, S., & Maddikunta, P. K. R. (2023). Cyberbullying detection solutions based on deep learning architectures. Multimedia Systems, 29(3), 1839-1852.

[6] Balaji, P. G., Katariya, P. P., Sruthi, S., & Venugopalan, M. (2024, April). Cyberbullying Detection on Multiclass Data Using Machine Learning and A Hybrid CNN-BiLSTM Architecture. In 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS) (Vol. 1, pp. 1-6). IEEE.

[7] Kumar, A. S., Kumar, N. S., Devi, R. K., & Muthukannan, M. (2024). Analysis of Deep Learning-Based Approaches for Spam Bots and Cyberbullying Detection in Online Social Networks. AI-Centric Modeling and Analytics, 324-361.

[8] Almomani, A., Nahar, K., Alauthman, M., Al-Betar, M. A., Yaseen, Q., & Gupta, B. B. (2024). Image cyberbullying detection and recognition using transfer deep machine learning. International Journal of Cognitive Computing in Engineering, 5, 14-26.

[9] Hasan, M. T., Hossain, M. A. E., Mukta, M. S. H., Akter, A., Ahmed, M., & Islam, S. (2023). A review on deep-learning-based cyberbullying detection. Future Internet, 15(5), 179.

[10] Akhter, A., Acharjee, U. K., Talukder, M. A., Islam, M. M., & Uddin, M. A. (2023). A robust hybrid machine learning model for Bengali cyber bullying detection in social media. Natural Language Processing Journal, 4, 100027.

[11] Paruchuri, V. L., & Rajesh, P. (2023). CyberNet: a hybrid deep CNN with N-gram feature selection for cyberbullying detection in online social networks. Evolutionary Intelligence, 16(6), 1935-1949.

[12] Suhas Bharadwaj, R., Kuzhalvaimozhi, S., & Vedavathi, N. (2022). A novel multimodal hybrid classifier based cyberbullying detection for

social media platform. In Proceedings of the Computational Methods in Systems and Software (pp. 689-699). Cham: Springer International Publishing.Murshed, B. A. H., Suresha, Abawajy, J., Saif, M. A. N., Abdulwahab, H. M., & Ghanem, F. A. (2023). FAEO-ECNN: cyberbullying detection in social media platforms using topic modelling and deep learning. Multimedia Tools and Applications, 82(30), 46611-46650.

[13] Lu, N., Wu, G., Zhang, Z., Zheng, Y., Ren, Y., & Choo, K. K. R. (2020). Cyberbullying detection in social media text based on character-level convolutional neural network with shortcuts. Concurrency and Computation: Practice and Experience, 32(23), e5627.

[14] Dutta, S., Neog, M., & Baruah, N. (2024, February). Towards Safer Social Spaces: LSTM, Bi-LSTM and Hybrid Approach for Cyberbullying Detection in Assamese language on Social Networks. In 2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE) (pp. 1-6). IEEE.

[15] Ellaky, Z., Benabbou, F., Matrane, Y., & Qaqa, S. (2024). A Hybrid Deep Learning Architecture for Social Media Bots Detection Based on Bigru-LSTM and Glove Word Embedding. IEEE Access.

[16] Omarov, B., & Altayeva, A. (2018, January). Towards intelligent IoT smart city platform based on OneM2M guideline: smart grid case study. In 2018 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 701-704). IEEE.

[17] Nitya Harshitha, T., Prabu, M., Suganya, E., Sountharrajan, S., Bavirisetti, D. P., Gadde, N., & Uppu, L. S. (2024). ProTect: a hybrid deep learning model for proactive detection of cyberbullying on social media. Frontiers in artificial intelligence, 7, 1269366.

[18] Roy, P. K., & Mali, F. U. (2022). Cyberbullying detection using deep transfer learning. Complex & Intelligent Systems, 8(6), 5449-5467.

[19] Shibly, F. H. A., Sharma, U., & Naleer, H. M. M. (2022). Detection of Cyberbullying in Social Media to Control Users' Mental Health Issues Using Recurrent Neural Network Architectures. Journal of Pharmaceutical Negative Results, 434-441.

[20] Daraghmi, E. Y., Qadan, S., Daraghmi, Y., Yussuf, R., Cheikhrouhou, O., & Baz, M. (2024). From Text to Insight: An Integrated CNN-BiLSTM-GRU Model for Arabic Cyberbullying Detection. IEEE Access.

[21] Nasution, M. A. S., & Setiawan, E. B. (2023). Enhancing Cyberbullying Detection on Indonesian Twitter: Leveraging Fast Text for Feature Expansion and Hybrid Approach Applying CNN and BiLSTM. Revue d'Intelligence Artificielle, 37(4), 929-936.

[22] Singh, N. M., & Sharma, S. K. (2024). An efficient automated multi-modal cyberbullying detection using decision fusion classifier on social media platforms. Multimedia Tools and Applications, 83(7), 20507-20535.

[23] Teng, T. H., & Varathan, K. D. (2023). Cyberbullying detection in social networks: A comparison between machine learning and transfer learning approaches. IEEE Access, 11, 55533-55560.

[24] Rodela, A. T., Nguyen, H. H., Farid, D. M., & Huda, M. N. (2023, October). Bangla Social Media Cyberbullying Detection Using Deep Learning. In International Conference on Intelligent Systems and Data Science (pp. 170-184). Singapore: Springer Nature Singapore.

[25] Omarov, B., Anarbayev, A., Turyskulov, U., Orazbayev, E., Erdenov, M., Ibrayev, A., & Kendzhaeva, B. (2020). Fuzzy-PID based self-adjusted indoor temperature control for ensuring thermal comfort in sport complexes. J. Theor. Appl. Inf. Technol, 98(11), 1-12.

[26] Pericherla, S., & Ilavarasan, E. (2024). Transformer network-based word embeddings approach for autonomous cyberbullying detection. International Journal of Intelligent Unmanned Systems, 12(1), 154-166.

[27] Jing, Y., Haowei, M., Ansari, A. S., Sucharitha, G., Omarov, B., Kumar, S., ... & Alyamani, K. A. (2023). Soft computing techniques for detecting cyberbullying in social multimedia data. ACM Journal of Data and Information Quality, 15(3), 1-14.

[28] Al Duhayyim, M., Mohamed, H. G., Alotaibi, S. S., Mahgoub, H., Mohamed, A., Motwakel, A., ... & Eldesouki, M. I. (2022). Hyperparameter Tuned Deep Learning Enabled Cyberbullying Classification in Social Media. Computers, Materials & Continua, 73(3).

[29] Omarov, B., Altayeva, A., Suleimenov, Z., Im Cho, Y., & Omarov, B. (2017, April). Design of fuzzy logic based controller for energy efficient operation in smart buildings. In 2017 First IEEE International Conference on Robotic Computing (IRC) (pp. 346-351). IEEE.

[30] Omarov, B. (2017, October). Development of fuzzy based smart building energy and comfort management system. In 2017 17th International Conference on Control, Automation and Systems (ICCAS) (pp. 400-405). IEEE.

[31] Obaid, M. H., Guirguis, S. K., & Elkaffas, S. M. (2023). Cyberbullying detection and severity determination model. IEEE Access.

[32] Al-Khasawneh, M. A., Faheem, M., Alarood, A. A., Habibullah, S., & Alsolami, E. (2024). Towards Multi-Modal Approach for Identification and Detection of Cyberbullying in Social Networks. IEEE Access.

[33] Ghosh, T., Chowdhury, A. A. K., Banna, M. H. A., Nahian, M. J. A., Kaiser, M. S., & Mahmud, M. (2022, October). A hybrid deep learning approach to detect bangla social media hate speech. In Proceedings of International Conference on Fourth Industrial Revolution and Beyond 2021 (pp. 711-722). Singapore: Springer Nature Singapore.

[34] Ahmed, M. T., Akter, N., Rahman, M., Das, D., & AZM T, R. G. (2023). Multimodal cyberbullying meme detection from social media using deep learning approach. Int J Comput Sci Inf Technol (IJCSIT), 15, 27-37.

[35] Kumari, K., Singh, J. P., Dwivedi, Y. K., & Rana, N. P. (2021). Multi-modal aggression identification using convolutional neural network and binary particle swarm optimization. Future Generation Computer Systems, 118, 187-197.

[36] Omarov, B. (2017, October). Exploring uncertainty of delays of the cloud-based web services. In 2017 17th International Conference on Control, Automation and Systems (ICCAS) (pp. 336-340). IEEE.

[37] Kadiri, P., Arjun, U., Sravani, N., Jyothi, N. S., Mahesh, P., & Naik, S. J. (2024, May). Detecting Cyberbullying through social media: A Deep Learning Approach. In 2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE) (pp. 1-7). IEEE.

[38] Altayeva, A. B., Omarov, B. S., Aitmagambetov, A. Z., Kendzhaeva, B. B., & Burkitbayeva, M. A. (2014). Modeling and exploring base station characteristics of LTE mobile networks. Life Science Journal, 11(6), 227-233.

[39] Bădiță, V. N. (2023). Cyberbullying Detection. International Journal of Information Security and Cybercrime (IJISC), 12(1), 37-44.

[40] Mahmud, T., Ptaszynski, M., Eronen, J., & Masui, F. (2023). Cyberbullying detection for low-resource languages and dialects: Review of the state of the art. Information Processing & Management, 60(5), 103454.

[41] Omarov, B., Orazbaev, E., Baimukhanbetov, B., Abusseitov, B., Khudiyarov, G., & Anarbayev, A. (2017). Test battery for comprehensive control in the training system of highly Skilled Wrestlers of Kazakhstan on national wrestling" Kazaksha Kuresi". Man In India, 97(11), 453-462.

[42] Akhter, M. P., Jiangbin, Z., Naqvi, I. R., AbdelMajeed, M., & Zia, T. (2022). Abusive language detection from social media comments using conventional machine learning and deep learning approaches. Multimedia Systems, 28(6), 1925-1940.

[43] Azzeh, M., Alhijawi, B., Tabbaza, A., Alabboshi, O., Hamdan, N., & Jaser, D. (2024). Arabic cyberbullying detection system using convolutional neural network and multi-head attention. International Journal of Speech Technology, 1-17.

[44] Omarov, B., Altayeva, A., & Cho, Y. I. (2017). Smart building climate control considering indoor and outdoor parameters. In Computer Information Systems and Industrial Management: 16th IFIP TC8 International Conference, CISIM 2017, Bialystok, Poland, June 16-18, 2017, Proceedings 16 (pp. 412-422). Springer International Publishing.

[45] Yi, P., & Zubiaga, A. (2023). Session-based cyberbullying detection in social media: A survey. Online Social Networks and Media, 36, 100250.

[46] Al-Wesabi, F. N., Obayya, M., Alabdan, R., Aljehane, N. O., Alazwari, S., Alruwaili, F. F., ... & Swathi, A. (2024). Automatic Recognition of Cyberbullying in the Web of Things and social media using Deep Learning Framework. IEEE Transactions on Big Data.

[47] Bhowmik, S., Sultana, S., Sajid, A. A., Reno, S., & Manjrekar, A. (2024). Robust multi-domain descriptive text classification leveraging conventional and hybrid deep learning models. International Journal of Information Technology, 16(5), 3219-3231.

Optimization with Adaptive Learning: A Better Approach for Reducing SSE to Fit Accurate Linear Regression Model for Prediction

Dr Vijay Kumar Verma[1], Dr Umesh Banodha[2], Dr Kamlesh Malpani[3]
Shri Vaishnav Vidyapeeth Vishwavidyalaya Indore, Computer Science and Engineering, Indore, M.P. India[1]
Dr. A.P.J. Abdul Kalam U.I.T. Jhabua, M.P. India[2]
Shri Vaishnav Institute of Management and Science, Indore M.P. India[3]

***Abstract*—The Optimization provides a way through which an optimum can be achieved. It is all about designing accurate and optimal output for a given problems with using minimum available resources. It is a task which refers to minimizing an objective function f(x) parameterized by x or it is the task which refers minimizing the cost function using the model's parameters. In machine learning optimization is slightly different. Usually most of the problems, are very much aware about shape, size and type of data. Such information helps us to know where need improve. In case of machine learning optimization works perfectly when there is no knowledge about new data. The method proposed in this paper is named as Optimization with adaptive learning which is used to minimize the cost in term of number of iterations for linear regression to fit the correct line for given dataset to reduce residual error. In regression analysis a curve or line fit in such a way for the data objects, that the differences of distances between the data points and curve or line is always minimum. Proposed approach Initialize random values for parameters of linear model and calculate Error (SSE). Our objective is minimizing the values of SSE, if SSE is large, need to adjust the selected initial values. The size of the step used in each iteration is direction movement to reach the local minimum for optimal value. After performing certain repetitions of the deviation, minimum value for SSE has found and it has a stable value with no change. Real life data set have been used for expositional analysis.**

***Keywords*—Adaptive learning; regression; optimization; minimum; cost; objective; error; random**

I. Introduction

Optimization is a mathematical technique used to solve quantitative problems in a number of discipline like physics, engineering, computer science etc. Several problems be formulated and solved by combining ideas and methods with the field of optimization. In computer Science mathematical programming contains the study of the mathematical optimization and the study of the mathematical properties of these approaches [21,18]. The advances development in the optimization techniques in computer science are used to solve number of problems in operations research, game theory, and numerical analysis. The problem used for optimization have three basic components shown in Fig. 1, objective function, variables, and set of constraints. The objectives are expected as a return of costs or profits. Variables are the quantities values and can be manipulated so that to objective function can be optimized. Set of constraints for an optimization problem are restricted for values which can choose for the variables or parameters [20]. Research problem that found is how to achieve optimization is linear regression. Adaptive learning provides a way to archives optimization by using random initialization of the parameters [13,15]. Objective of the proposed work is to achieve optimization and reduce SSE for accurate prediction. The proposed work not only helps to students but as help to training institution of JEE examination, to prediction the marks and possibility for selection.

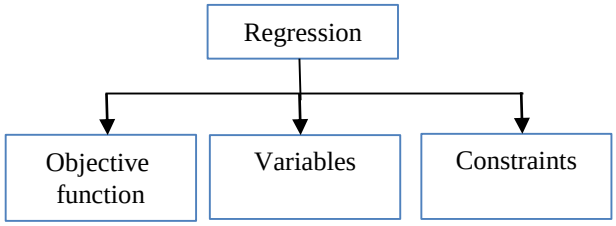


Fig. 1. Components used in regression analysis.

Regression analysis gives us number of benefits like

1) It shows important relations between variables.

2) It gives the effect of multiple explanatory variables.

II. Cost or Objective Function

The objective or cost function for analyzing regression is to find best fitted line for given inputs of x value and correct output for y value. Cost or objective function maps one or more variables on an actual instinctively representation related to predictor. ML models predict according to the new inputs [14,17].

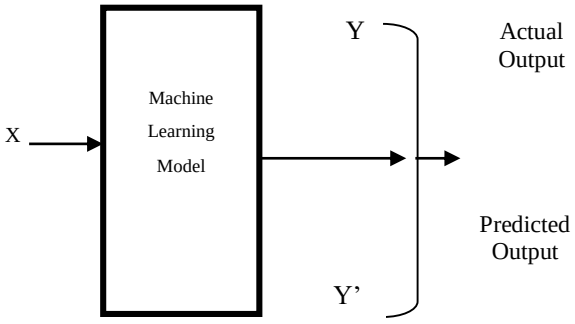


Fig. 2. Machine learning model.

The Error could be calculated by finding the difference between Predicted and Actual value:

Error = Y' (ML model Predicted value) - Y(Actual value)

A. Minimizing Error or Cost

Initial task of every ML model is to reduce cost. Minimizing value of cost functions will result to reduce error between the predicted value by ML model and actual value as shown in Fig. 2. Generally objective or cost function for linear regression is expressed in form of $\mathbf{Y = X^2}$. To minimize Objective or cost function, must find X to produce the correct Y value. The most common function required to minimize the parameters over a dataset is sum of squared error (SSE) and mean squared error (MSE). These errors ensure difference between the estimated value (prediction) and the estimator (the dataset) [16,19].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2$$

$$SSE = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

The above equation can adjust a little to make the calculation down the track a little simpler. Squared differences have used instead using absolute, this helps to derive a regression line [22]. To discover correct line, used first deviation for cost or objective function, and it is difficult to calculate deviation for absolute values as compared to squared values.

III. Problem with Existing System

Objective of regression analysis is to discover types of relation between response and predictor to get correct value for unknown dataset. It is difficult to find most appropriate regression line that attempts to predict and best fit line for given dataset for two points [16,19]. Number of methods have already exist, but the problems is how can optimize the solution Some time it is difficult to select following variables [20,23].

- Response Variable
- Predictor Variable(s)
- Stop
- Coefficients

1) What are the techniques used that the selected model better trained or learn the past relationship?

2) Predicted error should be minimized.

3) How Hyperplane is decided that minimizes the vertical offsets?

In machine learning terminology, optimization is the task of minimizing the cost or values of objective function, function by the model's parameters has following problems:

- Criteria are not fixed for deciding the initial value for parameter.
- Find the lowest possible value of the SSE to minimize the error.

IV. Objectives

For the proposed approach our objectives are:

1) Use random initialization of the pampers, to minimize error, to get optimal output for problem.

2) Continuously apply different updated values for the coefficients, estimate and select new coefficients that have a marginally better value with reduced error.

3) The objective is to deal with real life problems and dataset.

4) Use first order derivation on data for training and check performance and validation new data.

V. Review of Related Works

Several authors have been proposed their work to improve accuracy of prediction for regression analysis and published their work in various journals. In proposed work several related papers have been studied understand the concepts and basic things. Erasmus et al. of proposed method for establishing empirical relationships using linear regression. They describe that assumptions are comparatively poor, the LR model describing a restricted expectation. Specification for model must selected carefully and it is crucial when estimates for the coefficients of interest [1]. Shen Rong et al proposed regression model used in ML. They collected forecast temperature data for one year and the sale of iced products. They created a mathematical regression model which based on theory of data mining. The proposed Linear regression model is based on the practical situation. They implemented proposed model with the latest and most popular Python3.6 [2]. Katarina Valaskova et al proposed Financial Risk Measurement and Prediction Modelling. They used financial data of an enterprises and calculate important financial ratios which is responsible for health of the company. They used important predictors for forthcoming success. Multiple regression analysis where also used and identified the important predictors [3] Darman et al proposed Grade prediction with MLR for Mathematics. They predict the students' score with used of MLR for Final Exam. Score of students used as response variable in the model and the predictor variables are the assessment for tests. They a regression model with the adding SPSS tool [4]. Gaurav Pandeya et al. proposed RM model COVID-19 for. They have collected data for a particular time periods for year 2020. They evaluated performance using 1.75 for the regression model. This was helpful Government and doctors for preparation of their plans for the next two weeks [5]. K. K. Basee et al. proposed analyzing for various Regression Models. They used linear regression, multivariate and nonlinear regression models. They developed exponential, logistic type of regression. They used mat lab software and wrote a code without using pre-defined function [6]. Ira Sharma et al. proposed Linear Regression Model for factors Related with Carbon Stock. The objective is to evaluate the factors related with carbon stock. Data have been taken from department of Forest Research and Survey (DFRS). Linear regression showed a good fit of the model [7]. Samit Ghosal et al proposed Linear Regression to predict the number of deaths due to COVID-19 in India. They trace a trend related to death counts. They interchangeably applied Multiple and

linear regression analyses. They employed auto-regression technique to improve the prediction [8]. Yujiang et al. proposed An Adaptive Learning for Regression. They clarified problem and proposed a new methodological framework to forecast targets. The framework contains neural networks, to store newly collected data [9]. M Tanveer et al. proposed LR model for evaluating models to estimate stature. They applied LR and R2 values for preferred model. 100 persona data were used for experimental [10]. Dengyuhan Dai proposed Factors Affecting the Linear Regression Model for Data Analysis. They discussed detail effects of numerous parameters on the linear fit of randomly generated data sets. They include noise, independent variables, and different sample size. They minimized the effectiveness of noise. The optimal choice of each parameter can be obtained by comparing the goodness of fit [11]. Hai-Tao Jin, Fei Wang proposed Linear Regression Analysis of Sleep Quality. Data collected from various demographic area of the participants and psychological scales. They used sample *t*-test one-way ANOVA and showed comparisons [12].

VI. Proposed Approach

Step 1: To achieve rapid optimization, apply adaptive technique, which help to getting faster results with minimum iteration and less efforts. Initialize with random small value to the parameter. Let say parameters are a(c) and b(m) were

$$(Y_{pred} = m\ X+c)$$

a is constant and b slop parameter with random values and calculate error (SSE)

Step 2: For large SSE, apply first derivation and calculate new value for parameter a and b it has very small change from their original initialized value.

$$SSE = \frac{1}{2}(Y - YP)^2$$

This provides direction for movement for what value of a and

b with SSE should minimize.

Step 3: After calculating new value for parameter apply to reach the optimal solution and again calculate new SSE**.**

$$\partial Error/\partial c = -(Y_{actual} - Y_{predicted})$$

$$\partial Error/\partial m = -(Y_{actual} - Y_{predicted}) * X$$

Here value updating in SSE by

$$SSE = \frac{1}{2}(Y_{actual} - Y_{predicted})^2 = \frac{1}{2}(Y_{actual} - (m * X + c))^2$$

Steps 4: During training after specific iterations and make sure error rate must be decreases.

Step 5: Repeat steps 2 and 3 for continuously adjustments of parameters until doesn't significantly reduce the SSE.

Step 6: Sometimes can apply 2nd order derivative which are extremely fast and accurate.

VII. Illustrate with Example using Demo Dataset

TABLE I. Demo Dataset with 10 Records

S. No	No of Hours Study 15 days (120 hours)	Predicted Score (100)
1	10	12
2	23	22
3	25	58
4	34	20
5	36	55
6	45	39
7	44	54
8	58	53
9	94	99
10	101	61

Table I shows 10 sample records. Decimal scaling is used to make calculation easy. Table II shows result after applying decimal scaling.

TABLE II. Result of Decimal Scaling (normalization)

S.No	No of Hours Study per Month (210)	Predicted Score (100)
1	0.10	0.12
2	0.23	0.22
3	0.25	0.58
4	0.34	0.20
5	0.36	0.55
6	0.45	0.39
7	0.44	0.54
8	0.58	0.53
9	0.94	1.00
10	1.01	0.61

Step 1: To better fitted a line

$$Y_{pred} = mb\ X+c,$$

Start with random values for parameter c and m and compute forecast error (SSE). Let first small random values for parameter c and m

$$c=0.46 \text{ and } m=0.74$$

Step 2: Calculate the Error w.r.t the parameters

$$\partial Error/\partial c = -(Y_{actual} - Y_{predicted})$$

$$\partial Error/\partial m = -(Y_{actual} - Y_{predicted}) * X$$

Step 3: Now update value in SSE by

$$SSE = \frac{1}{2}(Y_{actual} - Y_{predicted})^2 = \frac{1}{2}(Y_{actual} - (m * X + c))^2$$

$\partial Error/\partial c$ *and* $\partial Error/\partial m$ give the direction of the movement of c, m w.r.t to Error. Table III shows the result of first derivation and total error.

TABLE III. CALCULATING SSE WITH RANDOM VALUE

Error	$\partial Error/\partial c$	$\partial Error/\partial m$
0.102	0.45	0.00
0.076	0.39	0.09
0.001	0.05	0.01
0.126	0.50	0.17
0.017	0.18	0.07
0.077	0.39	0.18
0.031	0.24	0.11
0.063	0.35	0.20
0.09	0.14	0.13
0.175	0.59	0.59
Total (Error)=0.668	3.210	1.455

TABLE IV. CALCULATING ERROR GRADIENT WITH RESPECT TO WEIGHT

$Y_{Predicted}$ = m X+c	Error	$\partial Error/\partial c$	$\partial Error/\partial m$
0.42	0.087	0.42	0.00
`0.58	0.064	0.36	0.08
0.59	0.000	0.01	0.00
0.66	0.107	0.46	0.15
0.69	0.010	0.14	0.05
0.74	0.063	0.36	0.16
0.74	0.021	0.20	0.09
0.84	0.048	0.31	0.18
1.10	0.005	0.10	0.09
1.15	0.148	0.54	0.54
	Total (Error)=0.543	2.830	1.410

Table IV shows the reduced error with respect to weight. It is found with random values of c and m, total Error=0.668. In this situation, need to update these values.

Step 4: Adjust value of parameter with the initial random value to reach the optimal values and minimized the error.

Now, need to update parameter of c and m so that move in the direction for optimization. Now need to calculate

$$c - \partial Error/\partial c$$

$$m - \partial Error/\partial m$$

Step 5: The adaptive rules are:

$$c = c - R * \partial Error/\partial c$$

$$m = m - R * \partial Error/\partial m$$

Here, is the Adaptive rate (R) have small value which can be adjustable. In this case adaptive rate value is 0.01.

Step 6: Now use updated value of c and m for prediction are c=0.43 and m=0.72. Now again calculate Error

It is clear that new prediction, the total Error has reduced form (0.668 to 0.543). Accuracy of the prediction has been improved.

Step 7: Repeat step 3 to 5 further adjustments in the value of c and m until significantly reduce the error. At a situation when no change in the error value stop the repetition at last, got the optimal and highest prediction accuracy.

Table V shows the reduced error with updated value.

TABLE V. CALCULATING ERROR WITH RESPECT TO UPDATED VALUE

S.No	**Predicted value**	**Error**
1	0.45	0.087
2	0.62	0.064
3	0.63	0.000
4	0.70	0.107
5	0.73	0.010
6	0.78	0.063
7	0.78	0.021
8	0.88	0.048
9	1.14	0.005
10	1.20	0.148
		Total= 0.553

VIII. EXPERIMENTAL ANALYSIS AND RESULT

A. Description Bout Dataset

Performance of proposed approach were evaluated with real life dataset. Dataset has been collected from 4 JEE entrance examination preparation institute Indore M.P. (ALLEN, KALPVIKSHA, FIITJEE, NARAYANA COACHING). More 1000 students' data have been collected. Datasets contain number of hours study and score during preparation. Based on the score in preparation for JEE they got selected in final JEE examination. Dataset used number of hours students' study in 15 days and appeared in JEE test series of 100 marks. Decimal scaling is used for number of hours. Python language is used to implementation and dataset stored in CSV data.

B. Description about Implementation

Python language to implement the roped approach, pandas and matplotlib library is used for implantation. Plot regression function have created and passed three parameters x, m and c, for unknown values to predict the correct value of y with high accuracy. In a new array predicted value are appended. Another function is created for updating error and update the value of m and c by using first derivation and calculate new value of error. Scatter plot is used to show number of hours study by the student and predicted score. Data points are very close so that they are overlapped. Fig. 3 shows the positive relationship between the variables.

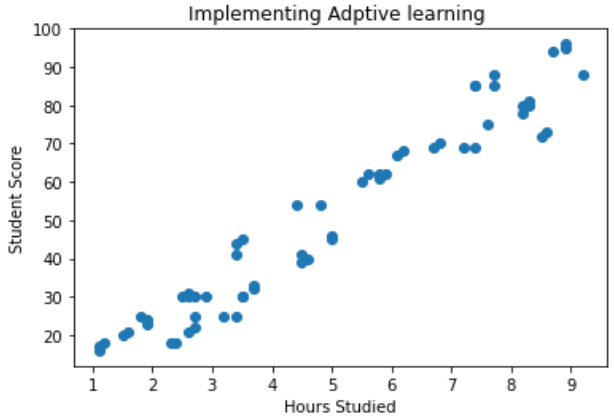


Fig. 3. Scatter plot for given dataset.

Initially the value selected for the m and c are 0.31 and 0.23, respectively. The selected valued used adaptive learning. From Fig. 4 staring with orange line and finally fitted the correct regression line for the given dataset. Fig. 4 shows the working process of adaptive learning. By the minimum number of iterations, got the best fitted line for the given dataset.

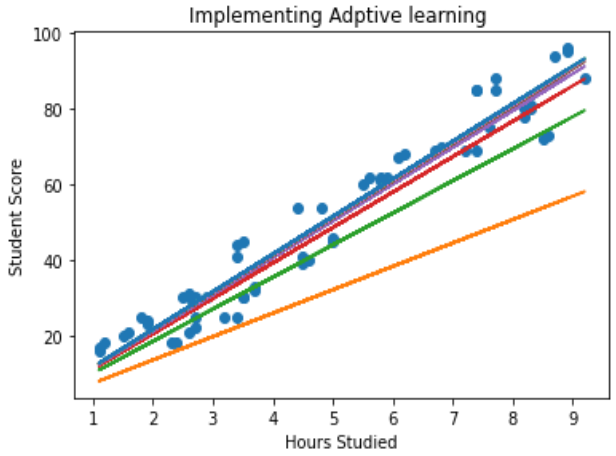


Fig. 4. Working process of adaptive learning.

In every iteration the reduced value of error has calculated. By the experimental analysis it is found that after 10 iteration the value of the error got minimized and has stable value (no minimization) at this point stop the process. Fig. 5 shows the change in value of parameters and the reduction in the error. Fig. 5 shows how the value of error gets reduced. In seconds iteration error is minimized to -5.33346109 just half of the first iteration. In the third iteration, it is minimized to -3.48410 and for fourth iteration it is minimized -2.762388. Similarly, in fifth, sixth, seventh, up to the tenth iteration the values of the error get minimized with very minor changes. So, after tenth iteration stop the process and got the best fitted line for given dataset.

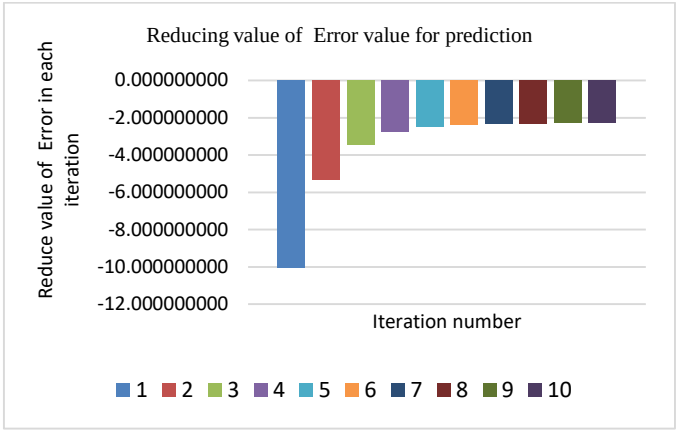


Fig. 5. Reducing error value in each iteration.

The value of the parameter m and c has been updated during each iteration. Fig. 6 shows the updated value of parameter m and c, from the Fig. 6 it is clear that after 5th iteration there are very minor changes are made in the value of m and c.

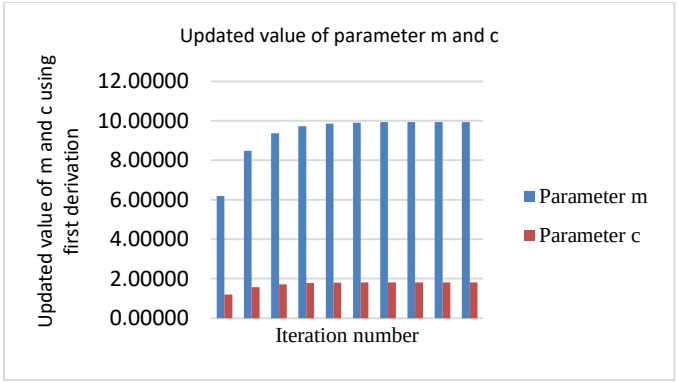


Fig. 6. Updating value of m and c during each iteration using first derivation.

So, prediction for unknown value of x the parameter has the value m=9.943 and c=1.816

C. Average prediction Error

Different number of records and calculated average prediction error have been used and found that Adaptive learning-based approach comparatively gave better prediction as other traditional approach. Table VI shows the prediction accuracy and average prediction error with number of students.

TABLE VI. PREDICTION ACCURACY AND AVERAGE PREDICTION ERROR

S No	No of Students	Prediction Accuracy	Average Prediction Error
1	890	96.31	3.69
2	2014	95.67	4.33
3	3022	97.23	2.77

Fig. 7 shows the graphical representation of number of students with prediction accuracy and average prediction error. It is also found that proposed approach is work well and proposed approach can scale up to any size of the dataset without effecting the performance.

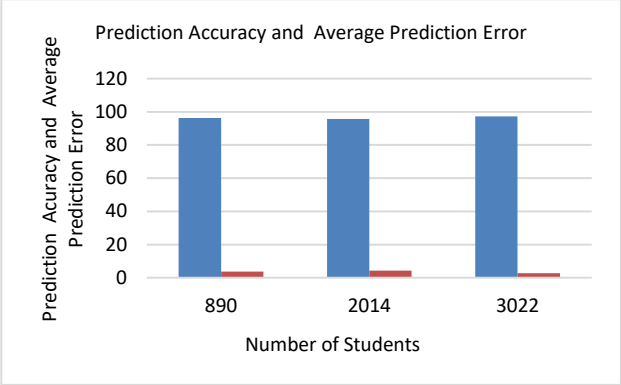


Fig. 7. Comparing prediction accuracy and average prediction error.

IX. Advantages and Limitations

The proposed approach has two advantages over the previous approach, it uses a minimum number of iterations and gives better accuracy with minimum error. The limitation of the proposed approach is that it uses only a single independent variable, there are several other parameters which can also be considered. The proposed approach is not suitable for nonlinear repression.

X. Conclusion and Future Work

Prediction is an important data analysis technique. Prediction helps us to predict what happens in future. In this paper, optimization with adaptive learning approach is proposed. Proposed approach improves accuracy of prediction by using adaptive learning and give optimal solution with minimum number of iterations. Real life data set is used for experimental analysis. Correct score can be predicted of the students based on number of studies for JEE exam. Dataset is collected from different JEE preparing centers of Indore city of Madhya Pradesh. By experimental analysis it is found that the accuracy of proposed approach for prediction is better and scalable. The proposed work helps students as well as training institution of JEE examination, to prediction the marks and possibility for selection. In future this approach can be applied for other type of regression and used for different datasets.

References

[1] Marno Verbeek Using linear regression to establish empirical relationships IZA World of Labor 2017: 336 doi: 10.15185/izawol.336 | Marno Verbeek © | February 2017 | wol.iza.org IZA World of Labor | February 2017 | wol.iza.org.

[2] Shen Rong Zhang Bao-wen The research of regression model in machine learning field MATEC Web of Conferences 176, 01033 (2018) https://doi.org/10.1051 2018 IFID 2018.

[3] Katarina Valaskova , Tomas Kliestik Financial Risk Measurement and Prediction Modelling for Sustainable Development of Business Entities Using Regression Analysis Sustainability 2018, 10, 2144; www.mdpi.com/journal/sustainability doi:10.3390/su10072144.

[4] Hazlina Darman, Sarah Musa Predicting Students' Final Grade in Mathematics Module using Multiple Linear Regression International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878 (Online), Volume-7 Issue-5s, January 2019.

[5] Gaurav Pandeya, Poonam Chaudhary SEIR and Regression Model based COVID-19 outbreak predictions in India https://doi.org/10.48550/arXiv.2004.00958 arXiv:2004.00958

[6] Baseer, K. K. and Neerugatti, Vikram and Tatekalva, Analysing Various Regression Models for Data Processing (June 30, 2019). International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-8 June, 2019.

[7] Ira Sharma and Sampurna Kakchapati Linear Regression Model to Identify the Factors Associated with Carbon Stock in Chure Forest of Nepal Hindawi Scientific Volume 2018, Article ID 1383482, 8 pages https://doi.org/10.1155/2018/1383482

[8] Samit Ghosal a Sumit Sengupta Linear Regression Analysis to predict the number of deaths in India due to SARS-CoV-2 at 6 weeks from day 0 (100 cases - March 14th 2020) Clinical Research & Reviews 14 (2020) 311e315 journal homepage: www.elsevier.com/locate/dsx

[9] Yujiang He and Bernhard Sick CLeaR: An Adaptive Continual Learning Framework for Regression Tasks arXiv:2101.00926v4 [cs.LG] 16 Jul 2021 Intelligent Embedded Systems (IES) Group, University of Kassel, Wilhelmshoher Allee 71 - 73, Kassel, Germany.

[10] M Tanveer Hossain Parash and Mohammad Mostafizur Simple linear regression approach for evaluating models to estimate stature based on upper limb dimensions of adult Bangladeshi Males Hossain Parash et al. Egyptian Journal of Forensic Sciences (2022) 12:20 https://doi.org/10.1186/s41935-022-00277-3.

[11] Dengyuhan Dai Several Factors Affecting the Linear Regression Model in Data Analysis ICMML 2023, November 24–26, 2023, Nanjing, China © 2023 ACM ISBN 979-8-4007-1697-3/23/11 https://doi.org/10.1145/3653724.3653734.

[12] Sharyn O'Halloran Linear Regression: A Model for the Mean Spring 2005

[13] D.S.G. POLLOCK: ECONOMETRICS The Classical Linear Regression Model basic theory of the classical statistical method of regression analysis.

[14] Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining Introduction to Linear Regression Analysis Fifth Edition.

[15] John O. Rawlings Sastry G. Pantula David A. Dickey Applied Regression Analysis: A Research Tool, Second Edition

[16] Robert Nau Fuqua Notes on linear regression analysis School of Business, Duke University (c) 2014 by Robert Nau, all rights reserved. Last updated on 11/26/2014.

[17] Walter A. Shewhart and SAMUEL S. WILKS Editors: Ruey S. Tsay, Introduction to Linear Regression Analys Wiley Series In Probability And Statistics Established By Sanford Weisberg Editors Emeriti: Vic Barnett, J. Stuart Hunter, Joseph B. Kadane, Jozef L. Teugel.

[18] Rebecca Bevans. Revised on July 17, 2020.An introduction to simple linear regression Published on February 19, 2020 .

[19] Julien I.E. Hoffman Variations Based on Linear Regression, in Biostatistics for Medical and Biomedical Practitioners, 2015.

[20] Jean-François Dupuy A Brief Overview of Linear Models, in Statistical Methods for Over dispersed Count Data, 2018.

[21] Claudia and Angelini Regression Analysis, in Encyclopedia of Bioinformatics and Computational Biology, 2019

[22] Astrid Schneider, Gerhard Hommel, and Maria Blettner "Linear Regression Analysis" Department of Medical Biometrics, Epidemiology, and Computer Sciences, Johannes Gutenberg University, Mainz, Germany.

[23] B. Van Schaeybroeck and S. Vannitsem "Post-processing through linear regression" Nonlin. Processes Geophys., 18, 147–160, 2011 doi:10.5194/npg-18-147-2011.

Efficient Remote Health Monitoring Using Deep Learning and Parallel Systems

Zakaria El Khadiri[1]*, Rachid Latif[2], Amine Saddik[3]
Laboratory of Systems Engineering and Information Technology (LISTI), National School of Applied Sciences, Ibn Zohr University Agadir, Agadir, Morocco[1, 2, 3]
Faculty of Applied Sciences, Ibn Zohr University, Ait Melloul, Morocco[3]

***Abstract*—This study presents a novel approach for non-contact extraction of physiological parameters, such as heart rate and respiratory rate, from facial images captured using RGB cameras, leveraging recent advancements in deep learning and signal processing techniques. The proposed system integrates Artifacts intelligent-driven algorithms for accurately estimating vital signs, addressing key challenges such as variations in lighting conditions, facial orientation, and noise. The system is implemented on both a naive homogeneous architecture and an optimized heterogeneous CPU-GPU system to enhance real-time performance and computational efficiency. A comparative analysis is performed to evaluate processing speed, accuracy, and resource utilization across both architectures. Results demonstrate that the optimized heterogeneous system significantly outperforms the homogeneous counterpart, achieving faster processing times while maintaining high accuracy levels. This performance boost is achieved through parallel computing frameworks such as OpenMP and OpenCL, which allow for efficient resource allocation and scalability. The research highlights the potential of heterogeneous architectures for real-time healthcare applications, including remote patient monitoring and telemedicine, providing a robust solution for future developments in non-invasive health technology.**

***Keywords*—Real-time healthcare; embedded systems; heterogeneous computing; deep learning; CPU-GPU architecture**

I. Introduction

A. Background and Motivation

In recent years, the healthcare industry has experienced a significant shift towards non-contact monitoring solutions, driven by the increasing demand for continuous and unobtrusive patient care. Traditional contact-based physiological monitoring methods, such as electrocardiograms (ECGs) and wearable biosensors, while effective in providing accurate measurements, often face several practical challenges. These challenges include patient discomfort, the need for frequent repositioning of sensors, and hygiene concerns, which can limit patient compliance and the frequency of monitoring. Moreover, the inconvenience of attaching and removing sensors can be a barrier to widespread adoption, particularly in settings that require long-term or continuous monitoring.

In response to these limitations, researchers and engineers have turned their attention to non-contact methods, specifically those leveraging facial image analysis. By utilizing standard RGB cameras, which are less intrusive and can be easily integrated into everyday environments, it is possible to monitor physiological parameters such as heart rate and respiratory rate without direct physical contact. This non-contact approach not only enhances patient comfort but also facilitates continuous and real-time monitoring, allowing for more comprehensive health assessments over time [1]-[5].

Recent advancements in deep learning and signal processing have greatly enhanced the feasibility and accuracy of extracting physiological parameters from facial images. Deep learning algorithms, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have demonstrated remarkable capabilities in interpreting complex patterns within facial images, even under varying lighting conditions and different facial orientations. Signal processing techniques further refine these interpretations by analyzing subtle color changes and motion artifacts associated with physiological processes [6]-[10].

The integration of these advanced technologies into non-contact monitoring systems presents a transformative opportunity for the healthcare sector. Such systems promise not only to improve patient comfort and compliance but also to expand the reach of remote monitoring, making it possible to deliver continuous care in a variety of settings, including home environments and telemedicine platforms. This shift towards more seamless and less intrusive monitoring aligns with broader trends in healthcare innovation, aiming to enhance patient outcomes through more accessible, real-time, and data-driven approaches. As the field evolves, the development of robust, efficient, and accurate non-contact monitoring systems will play a crucial role in shaping the future of healthcare, offering new possibilities for early detection, preventive care, and personalized treatment.

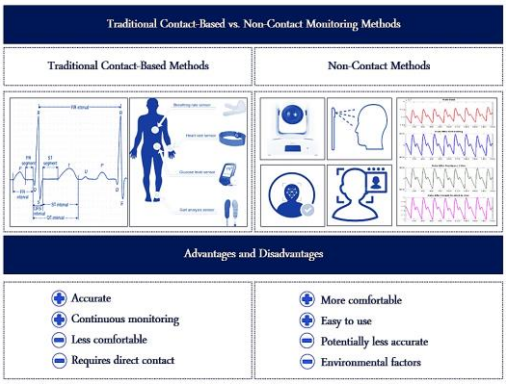


Fig. 1. Traditional contact-based vs. non-contact monitoring methods.

*Corresponding Author

In the above Fig. 1, a diagram comparing traditional contact-based physiological monitoring methods (e.g., ECG, wearable sensors) with non-contact methods (e.g., facial image analysis). The diagram highlights advantages such as comfort and ease of use for non-contact methods.

B. Advances in Deep Learning and Signal Processing

Recent advances in deep learning and signal processing have brought transformative improvements to the field of non-contact physiological monitoring, particularly in analyzing facial images to extract vital signs such as heart rate and respiratory rate. Deep learning techniques, especially convolutional neural networks (CNNs), have proven to be highly effective in analyzing complex visual data. CNNs are particularly adept at detecting subtle, pixel-level changes in facial images, such as variations in skin tone caused by the underlying blood flow, a method commonly referred to as remote photoplethysmography (rPPG). These subtle changes, which are often invisible to the human eye, correlate directly with physiological parameters like heart rate, enabling precise and continuous monitoring without the need for physical contact [11]-[13].

Further advances in recurrent neural networks (RNNs), including long short-term memory networks (LSTMs), have enhanced the capability to model temporal dependencies and sequences, making them well-suited for tracking cyclic patterns such as respiratory rate. By utilizing temporal data from video sequences, RNNs can capture periodic facial movements corresponding to breathing patterns, allowing for the accurate estimation of respiratory rates. This combination of CNNs for spatial feature extraction and RNNs for temporal analysis creates a powerful framework for real-time, non-invasive physiological monitoring.

On the signal processing front, techniques such as optical flow, which detects movement by calculating changes in pixel intensities between consecutive frames, and discrete wavelet transforms (DWT), which decompose signals into multi-resolution components, have further optimized the extraction of physiological signals. These methods work in tandem with deep learning models, refining the input data and enhancing the accuracy of parameter estimation. Additionally, advanced filtering algorithms, such as bandpass filters, are often employed to remove noise and isolate the relevant physiological signal, especially under challenging conditions like varying lighting, motion artifacts, and changes in facial orientation [14]-[16].

The integration of these deep learning and signal processing techniques with heterogeneous computing platforms, such as CPU-GPU architectures, has significantly improved system performance. By distributing computational tasks across multiple processing units, such systems offer enhanced scalability, reduced latency, and faster real-time processing capabilities. This has profound implications for applications in remote health monitoring, where real-time accuracy and computational efficiency are critical. The continued refinement of these technologies promises to further elevate the feasibility of non-contact physiological monitoring for widespread use in telemedicine, smart health environments, and continuous remote patient care [17]-[20]. In the following Fig. 2, a flowchart illustrating the workflow of deep learning and signal processing techniques used in facial image analysis. The flowchart includes steps such as image acquisition, preprocessing, feature extraction, and parameter estimation.

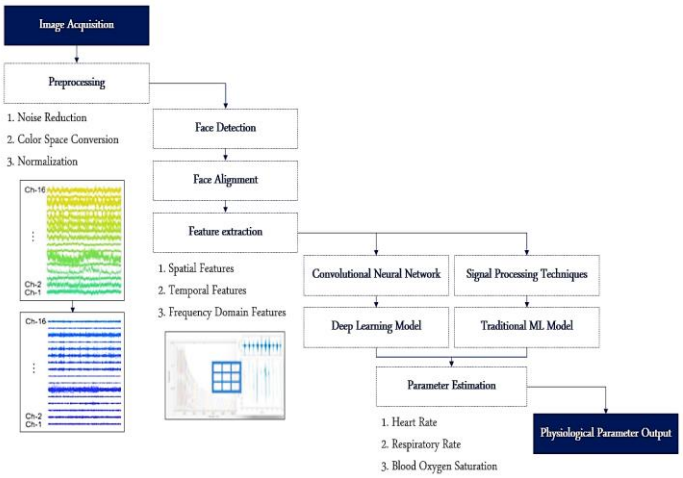


Fig. 2. Deep learning and signal processing workflow.

C. Contributions

In recent years, notable progress has been made in non-contact physiological monitoring using RGB cameras. However, many existing methods still face significant challenges. These include reduced accuracy in dynamic lighting conditions, handling diverse facial orientations, and noise in the captured signals. Additionally, although deep learning and signal processing techniques have been integrated into some systems, the real-time performance is often hindered by the high computational demands, particularly when running on homogeneous architectures such as CPU-only platforms. This creates limitations that make it difficult to deploy these systems in practical, real-time healthcare settings.

To overcome these challenges, our study introduces an optimized approach that utilizes heterogeneous architectures, specifically leveraging CPU-GPU systems, for real-time extraction of physiological parameters from facial images. This approach enhances processing speed without compromising accuracy. Furthermore, we employ advanced AI algorithms tailored to minimize the impact of noise, variations in lighting, and changes in facial orientation, thereby increasing the overall reliability and robustness of the system.

The overarching goal of this research is to design and implement a robust, non-contact system for accurately extracting physiological parameters, such as heart rate and respiratory rate, from facial images captured by RGB cameras. This system leverages state-of-the-art deep learning models and advanced signal processing techniques to address the inherent challenges in non-contact monitoring, including variations in facial orientation, changes in lighting conditions, and noise introduced by environmental factors. One of the primary contributions of this study is the development of a novel AI-driven framework that integrates convolutional neural networks (CNNs) for feature extraction with signal processing algorithms to analyze the subtle changes in facial color and movement, which correlate with vital signs. Furthermore, the

system is designed to be optimized for real-time performance by implementing it on a heterogeneous CPU-GPU architecture. This architecture enables parallel processing and efficient resource allocation, thus significantly enhancing computational efficiency. A comparative analysis is conducted between the proposed optimized heterogeneous implementation and a naive homogeneous system to evaluate improvements in processing speed, and accuracy. The research highlights that the optimized heterogeneous system achieves superior performance in real-time applications, making it well-suited for critical healthcare scenarios such as remote patient monitoring and telemedicine. Additionally, this study provides valuable insights into the advantages of parallel computing frameworks, such as OpenMP and OpenCL, in optimizing the execution of deep learning and signal processing algorithms, contributing to the broader field of non-invasive health monitoring technologies. Moreover, the following Fig. 3 is a diagram showing the overall system architecture, from facial image acquisition to physiological parameter extraction. The diagram highlights components such as the camera, processing unit, deep learning model, and output analysis.

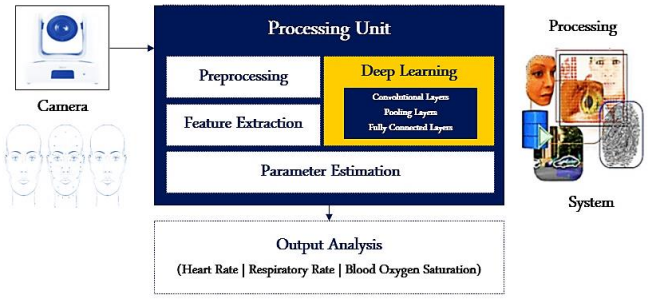


Fig. 3. System architecture and workflow.

II. State-of-the-Art: Review

A. Physiological Parameter Extraction from Facial Features

Recent advancements in non-contact physiological monitoring have enabled the extraction of vital signs, such as heart rate (HR), respiratory rate (RR), and blood oxygen levels, directly from facial images using RGB cameras. This approach leverages subtle physiological cues, primarily through techniques like Remote Photoplethysmography (rPPG), which detects minute changes in skin color caused by blood flow under the skin's surface [21]-[26]. These variations are captured as pixel intensity changes, invisible to the naked eye but detectable by advanced image processing algorithms. rPPG-based methods rely on capturing video streams of the subject's face and analyzing the temporal patterns of these pixel changes to estimate the heart rate. Similarly, respiratory rate estimation often employs optical flow algorithms that track small chest and shoulder movements associated with breathing, translating pixel displacements over time into respiration patterns. While highly effective, these techniques are sensitive to various factors such as lighting conditions, head motion, and camera quality. To mitigate these challenges, recent studies have incorporated advanced machine learning algorithms to enhance robustness, allowing for more accurate physiological measurements in real time, even under suboptimal conditions. This non-invasive approach holds tremendous promise for healthcare applications, enabling continuous monitoring without the need for physical contact, making it particularly useful for remote patient monitoring and telemedicine [27]-[33].

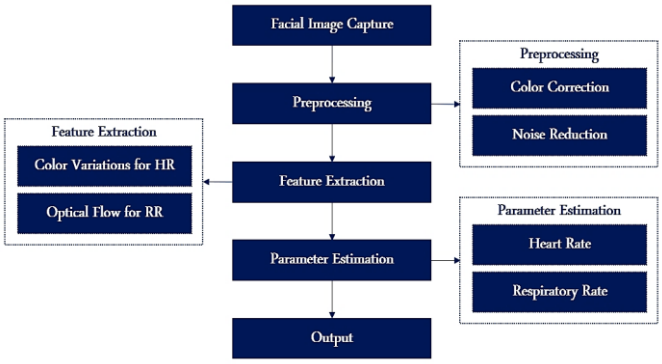


Fig. 4. Physiological parameter extraction process.

In Fig. 4, a flowchart depicting the physiological parameter extraction process, the facial image capture initiates the process, followed by preprocessing steps such as color correction and noise reduction to enhance image quality. Next, feature extraction is performed, where color variations are analyzed to estimate heart rate (HR), and optical flow techniques are used to detect respiratory rate (RR). These extracted features are then processed in the parameter estimation phase to calculate accurate values for both heart rate and respiratory rate. Finally, the system outputs these physiological parameters, providing non-contact monitoring results.

B. Image and Signal Processing Techniques

Image and signal processing techniques are fundamental in extracting physiological parameters from facial images, addressing challenges related to accuracy, robustness, and environmental variability. One of the primary techniques involves color space conversion, where facial images captured in the RGB format are transformed into alternative color spaces, such as YUV, YCbCr, or HSV. These transformations are critical as they allow the separation of luminance and chrominance components, which significantly improves the detection of subtle color variations in the skin caused by blood flow, a key indicator for heart rate estimation. Specifically, the chrominance channels (U and V in YUV, or Cb and Cr in YCbCr) are more sensitive to these physiological changes, making them ideal for accurate feature extraction [34]-[36].

An essential step in this process is skin detection, where the region of interest (ROI) is isolated to ensure that only skin pixels are analyzed. Several skin detection algorithms are employed, ranging from traditional thresholding techniques based on predefined color ranges to more sophisticated methods using machine learning models like support vector machines (SVMs) and neural networks (NN). These advanced methods adaptively classify skin regions based on training data, making the system more robust to variations in lighting and individual skin tones. Additionally, motion artifact reduction is a crucial aspect, as head movements and facial expressions can introduce noise into the signal. Techniques

such as optical flow analysis are used to track pixel displacement over time, isolating and compensating for movements unrelated to physiological signals.

In parallel, advanced signal processing methods are applied to the extracted physiological signals to enhance their clarity and improve the accuracy of parameter estimation. One widely used approach is the wavelet transform, which provides a multi-resolution analysis of the signal, capturing both time and frequency information. This method is particularly effective in isolating the periodic components corresponding to physiological processes, such as heartbeats or respiration, from background noise. Similarly, Fourier analysis is employed to transform the time-domain signal into the frequency domain, where periodic features, such as the heart rate frequency, can be more easily identified. These techniques are instrumental in filtering out high-frequency noise or low-frequency drifts that could otherwise distort the signal.

Moreover, motion-compensating filtering algorithms are often integrated into the processing pipeline to mitigate artifacts caused by slight movements of the face or background disturbances. Combined with the signal processing techniques mentioned earlier, these algorithms enhance the system's ability to produce reliable physiological parameter estimates in real-time. Overall, the synergy between image processing (such as color space conversion and skin detection) and signal processing (like wavelet transform and Fourier analysis) enables a more robust extraction of physiological parameters, ensuring that the system can operate effectively under varying conditions, such as fluctuating lighting, diverse skin tones, and slight motion disturbances. These advancements are pivotal in making non-contact physiological monitoring systems practical for real-world healthcare applications, including remote monitoring and telemedicine.

C. AI and Deep Learning in Physiological Monitoring

The incorporation of artificial intelligence (AI) and deep learning techniques has revolutionized the field of physiological monitoring from facial images, addressing key challenges such as environmental variability, facial orientation, and signal noise. Convolutional Neural Networks (CNNs) are at the forefront of feature extraction from facial images, utilizing multiple layers of convolutions to identify and learn intricate patterns related to physiological signals. CNNs excel in detecting subtle color variations and spatial features in facial skin that correlate with blood flow, enabling accurate heart rate (HR) and blood oxygen level estimation. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, are employed to capture temporal dependencies in the data, which is crucial for analyzing dynamic physiological signals like respiratory rate (RR). LSTMs are designed to handle sequential data by retaining information over long periods, making them adept at predicting and analyzing continuous physiological variations. Furthermore, Generative Adversarial Networks (GANs) are increasingly utilized to address data quality issues by generating synthetic data that resembles real physiological signals. GANs improve signal robustness by learning from real signal distributions and correcting distortions or noise, thus enhancing the accuracy of measurements even in challenging conditions. Together, these AI-driven approaches enable sophisticated real-time physiological monitoring by providing high precision in signal extraction and analysis. The deployment of these advanced algorithms not only enhances the accuracy and reliability of non-contact health monitoring systems but also paves the way for more effective telemedicine solutions and remote patient care, showcasing the transformative impact of AI and deep learning on healthcare technology.

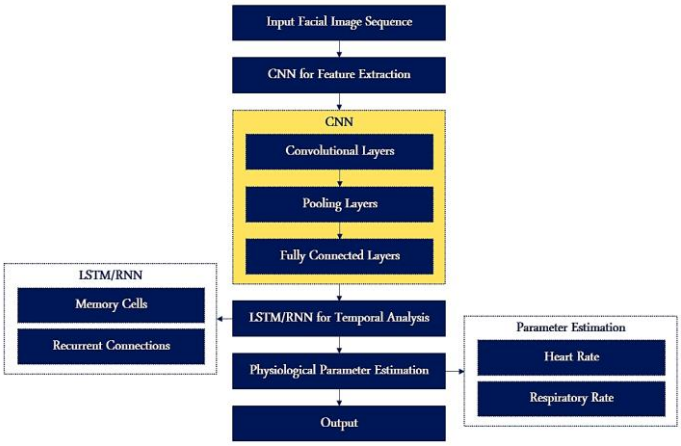


Fig. 5. AI-driven physiological monitoring.

In Fig. 5, the process begins with the input of a facial image sequence, which is then passed through a Convolutional Neural Network (CNN) for feature extraction to identify key physiological indicators from the image data. Following this, a Long Short-Term Memory (LSTM) or Recurrent Neural Network (RNN) performs temporal analysis to capture time-dependent changes in the extracted features. These changes are used to estimate physiological parameters, such as heart rate (HR) and respiratory rate (RR). The final step is the output, where the system provides the estimated physiological parameters for real-time monitoring.

D. Hardware Implementations for Physiological Monitoring

The hardware implementations for physiological monitoring systems have evolved substantially, driven by the need for real-time processing and enhanced computational capabilities. Historically, homogeneous systems relying solely on Central Processing Units (CPUs) provided a foundational approach to processing physiological data. While effective for basic tasks, these systems face limitations in handling the high computational demands associated with advanced image and signal processing algorithms, particularly those involving deep learning techniques. The limitations of homogeneous systems are primarily related to their restricted ability to perform parallel computations, which are crucial for real-time analysis of large-scale data such as facial images [37]-[39].

To overcome these constraints, heterogeneous systems have emerged, incorporating both CPUs and Graphics Processing Units (GPUs). These systems leverage the parallel processing capabilities of GPUs to handle intensive computational tasks more efficiently than traditional CPUs. For example, platforms such as the Odroid XU4 integrate a high-performance ARM Cortex-A15 CPU with an ARM Mali-T628 GPU. This combination allows for the simultaneous execution

of multiple processing threads, significantly accelerating tasks such as image pre-processing, feature extraction, and physiological parameter estimation.

The integration of heterogeneous architectures enables the utilization of parallel computing frameworks like OpenMP and OpenCL. OpenMP facilitates the efficient execution of multi-threaded applications by allowing developers to parallelize code across multiple CPU cores, thus enhancing the performance of data-intensive tasks. OpenCL, on the other hand, extends this parallelism to GPU cores, offering a robust environment for executing complex algorithms related to image and signal processing. By distributing processing workloads between CPUs and GPUs, heterogeneous systems can achieve substantial improvements in processing speed and real-time performance, which are critical for applications in remote physiological monitoring and telemedicine (Fig. 6) [40]-[42].

Despite the advantages of heterogeneous systems, several challenges persist. Optimizing these systems for energy efficiency remains a key concern, especially in mobile and embedded applications where power consumption is a critical factor. Additionally, the scalability of heterogeneous systems poses challenges as the complexity of algorithms and the volume of data increase. Future developments in hardware architecture and optimization techniques are essential to address these challenges and enhance the practicality of heterogeneous systems for broader applications in non-contact physiological monitoring.

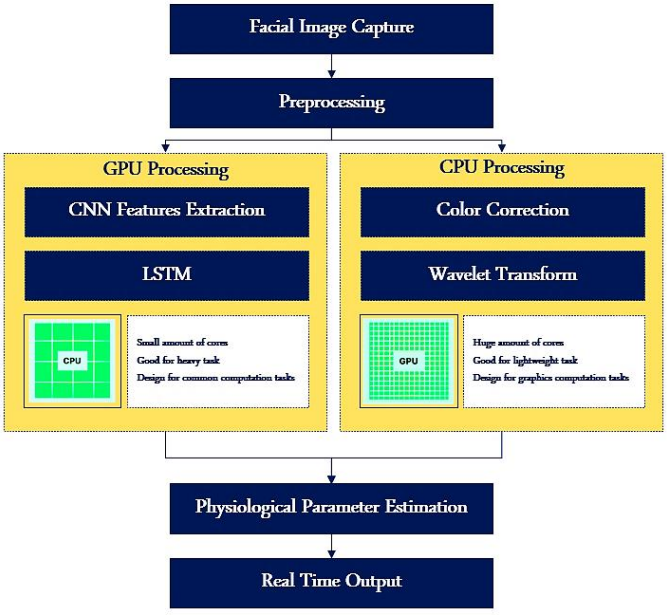


Fig. 6. Heterogeneous system architecture for physiological monitoring.

III. Proposed Methodology

A. Image and Signal Processing Algorithm

The image and signal processing algorithm for extracting physiological parameters from facial images is a multifaceted approach that integrates advanced image processing and signal analysis techniques to achieve high accuracy and robustness. Initially, high-resolution facial images are captured using a high-definition RGB camera, with the setup optimized to minimize variations in lighting conditions and facial orientation, ensuring consistent and reliable image data. The preprocessing phase involves several critical steps: first, face detection algorithms, such as Haar cascades or Multi-task Cascaded Convolutional Networks (MTCNN), are employed to precisely locate and extract the facial region from the acquired images. Subsequent normalization processes are applied to standardize image dimensions and correct for color imbalances, thereby mitigating the effects of external variables. The extracted Region of Interest (ROI) within the facial area is then subjected to detailed color analysis to detect minute variations in skin tone, which are indicative of changes in blood flow. These color fluctuations are converted into Photoplethysmographic (PPG) signals through specialized signal extraction techniques. The PPG signals, which reflect periodic variations in blood volume, are analyzed to determine heart rate and respiratory rate. Heart rate estimation is performed using Fourier Transform methods, such as Fast Fourier Transform (FFT), to identify dominant frequency components associated with cardiac activity. Respiratory rate estimation is achieved by analyzing the amplitude and frequency variations in the PPG signals, which correspond to respiratory cycles. This comprehensive algorithmic approach ensures the accurate and real-time extraction of physiological parameters, providing a solid foundation for subsequent applications in remote health monitoring and telemedicine. By combining these advanced image and signal processing techniques, the system is capable of delivering precise and actionable health insights from non-contact facial imaging.

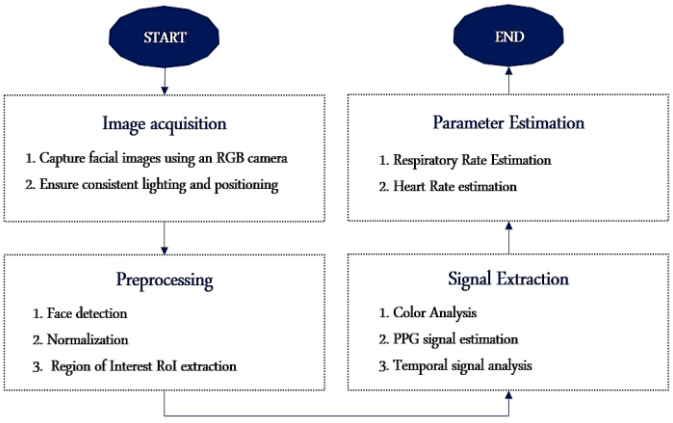


Fig. 7. Flowchart of the image and signal processing algorithm.

In Fig. 7, the image and signal processing algorithm for extracting physiological parameters from facial images involves a series of crucial steps. First, facial images are acquired using an RGB camera, capturing the necessary visual data. Next, preprocessing techniques are applied to enhance image quality and extract relevant features, such as skin tone variations. In the signal extraction phase, the algorithm analyzes subtle facial color variations to infer underlying physiological signals. Finally, parameter estimation methods are employed to process these signals and accurately derive heart rate and respiratory rate, providing non-invasive health monitoring results.

B. Deep Learning Component

The deep learning component of the proposed system is designed to significantly enhance the extraction and estimation of physiological parameters from facial images through the application of sophisticated neural network architectures. At the core of this component is the use of Convolutional Neural Networks (CNNs), which are instrumental in performing feature extraction from the raw facial images (Fig. 8). The CNN architecture comprises multiple layers, including convolutional layers that apply a series of filters to the input images to detect fundamental features such as edges and textures. These are followed by pooling layers, which reduce the dimensionality of the feature maps while preserving essential information. This hierarchical feature extraction enables the network to capture both low-level and high-level facial attributes pertinent to physiological signal analysis.

The processed features are then fed into a Recurrent Neural Network (RNN) component, specifically Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU), to address the temporal dynamics inherent in the physiological signals (Fig. 8). The RNNs are adept at capturing temporal dependencies and patterns within the time-series data extracted from the facial images, such as the periodic fluctuations in blood flow related to heart rate and respiration. By modeling these temporal relationships, the network improves the accuracy and robustness of the physiological parameter estimates.

The training of these neural networks involves using a large and diverse dataset of labeled facial images with known physiological parameters. The loss function, typically Mean Squared Error (MSE) for regression tasks, measures the discrepancy between the predicted and actual parameter values. Optimization algorithms such as Adam or Stochastic Gradient Descent (SGD) are employed to minimize this loss function by adjusting the network's weights and biases iteratively. This training process ensures that the network learns to generalize well across different individuals and conditions, enhancing its performance in real-world applications.

Furthermore, integrating these deep learning models with the signal processing pipeline is crucial for achieving real-time performance. The system benefits from parallel processing frameworks such as OpenMP and OpenCL, which are employed to optimize computational efficiency and reduce processing latency. This combined approach not only facilitates accurate and timely extraction of heart rate and respiratory rate from facial images but also ensures the system's scalability and adaptability to various deployment scenarios, including remote health monitoring and telemedicine applications. As shown in Fig. 8, the system integrates a Convolutional Neural Network (CNN) for feature extraction from facial images, utilizing convolutional layers to identify key features, pooling layers to reduce dimensionality while retaining crucial information, and fully connected layers to interpret these features for prediction. A Recurrent Neural Network (RNN), equipped with LSTM or GRU layers, processes time-series data from PPG signals to capture temporal dependencies, with an output layer predicting physiological parameters.

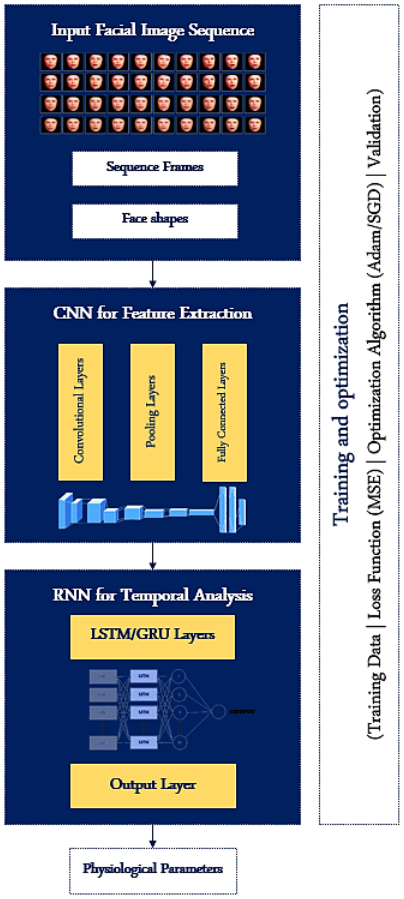


Fig. 8. Neural network architecture for physiological parameter extraction.

C. Deep Learning and AI Models

The global system architecture is meticulously designed to facilitate the efficient extraction of physiological parameters from facial images through a series of integrated modules, each performing a critical function within the overall framework. The architecture begins with the Image Acquisition Module, which utilizes high-resolution RGB cameras to capture continuous or periodic facial images under controlled lighting conditions to ensure image consistency. These images are then processed by the Preprocessing Module, which encompasses several key operations: face detection using algorithms such as Haar cascades or MTCNN, normalization of image size and color balance to mitigate variability, and extraction of the Region of Interest (ROI) where physiological signals are most prominent.

Following preprocessing, the Signal Processing Module analyzes the facial color variations within the ROI to extract photoplethysmographic (PPG) signals. This step involves sophisticated techniques to detect subtle changes in skin color due to blood flow, which are then used to derive the heart rate and respiratory rate. The extracted signals are subjected to temporal analysis to enhance accuracy.

The core of the system's analytical capabilities resides in the AI Module, which applies advanced deep learning techniques to process the extracted signals. This module incorporates Convolutional Neural Networks (CNNs) for

feature extraction from the images and Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRUs), for analyzing temporal dependencies in the PPG signals. The Output Module consolidates the results, providing a user-friendly interface for displaying or transmitting the estimated physiological parameters. This module ensures that the data is presented in a format suitable for further analysis or integration with healthcare systems.

The system is implemented on both a naive homogeneous architecture and an optimized heterogeneous architecture. The homogeneous system relies solely on CPU resources, potentially limiting processing speed and efficiency. In contrast, the heterogeneous system harnesses both CPU and GPU capabilities, employing parallel computing frameworks such as OpenMP and OpenCL to enhance processing performance. The heterogeneous system allows for scalable and efficient resource allocation, significantly improving real-time processing capabilities and making it highly suitable for practical healthcare applications such as remote monitoring and telemedicine.

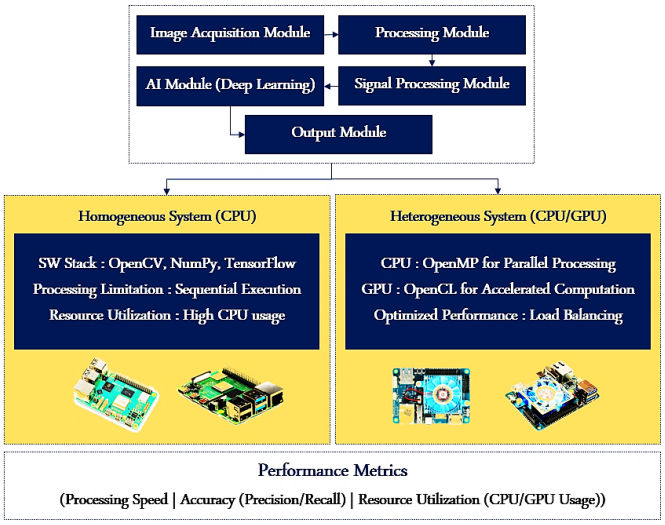


Fig. 9. System architecture diagram.

In Fig. 9 above, the system consists of several interconnected components. The image acquisition module captures facial images and sends them to the preprocessing module, which enhances image quality and extracts the region of interest (ROI) for further analysis. The signal processing module then examines color changes in the facial images, extracting photoplethysmography (PPG) signals. The AI module applies deep learning models to these signals to estimate physiological parameters such as heart rate and respiratory rate. Finally, the output module displays or transmits the extracted parameters for real-time monitoring or further processing.

IV. System Implementation

A. Homogeneous System

In the homogeneous system implementation, all computational tasks are executed on a single Central Processing Unit (CPU), which manages the entire workflow of facial image analysis for physiological parameter extraction. The process begins with the acquisition of RGB images through a standard camera setup. These images are subjected to a series of preprocessing steps to ensure uniformity and accuracy in subsequent analysis. The preprocessing phase includes facial detection using established algorithms such as Haar Cascades or Multi-task Cascaded Convolutional Networks (MTCNN). These algorithms identify and locate facial regions within the captured images, which are then cropped and normalized to mitigate variations in lighting, scale, and orientation.

Once the facial regions are isolated, the system employs remote photoplethysmography (rPPG) techniques to extract temporal signals associated with physiological parameters from these facial areas. rPPG relies on subtle variations in facial skin color that correspond to cardiovascular changes, which are indicative of heart rate and respiratory rate. Feature extraction is carried out using Convolutional Neural Networks (CNNs), which are trained to recognize patterns in the temporal signals and extract relevant features indicative of physiological states.

The extracted features are then processed through a series of signal-processing algorithms to estimate physiological parameters. This involves computing heart rate and respiratory rate from the temporal signals, with additional post-processing steps to filter out the noise and smooth the data for accurate parameter estimation. Despite its functional capability, the homogeneous system's reliance on a single CPU for all processing tasks poses constraints in terms of processing speed and real-time performance. This limitation is particularly evident when dealing with high-resolution images or when requiring rapid processing to meet real-time monitoring demands. The system's performance may be hindered by the CPU's inability to efficiently handle the computational load and parallelize tasks, leading to potential delays in parameter extraction and analysis (Table I).

TABLE I. Homogeneous System Implementation

Function	Facial image acquisition	Preprocessing	Feature extraction	Parameter estimation
Sub-Function	Capture RGB images	Face detection	Signal extraction rPPG	Signal processing
	Preprocess images	Image normalization	DL processing CNN	Post-processing
Output	Sequence of images	Normalized facial ROIs	Extracted signals & features	Estimated vital signs

B. Heterogeneous System

The optimized version of the system leverages a heterogeneous architecture, which integrates a multi-core CPU with a dedicated GPU to achieve significant improvements in both computational speed and efficiency. In this architecture, the computational tasks are strategically partitioned between the CPU and GPU to maximize resource utilization and minimize processing time. The initial steps, such as facial image acquisition and preprocessing (including face detection and normalization), are handled by the CPU. This ensures that simpler, less resource-intensive tasks are managed by the CPU,

freeing the GPU for more computationally demanding operations. Following the detection of the region of interest (ROI) in the facial image, the system offloads the critical task of signal extraction using remote photoplethysmography (rPPG) to the GPU. The GPU, with its parallel processing capabilities, efficiently handles the large datasets and intensive computations required for rPPG signal extraction and subsequent feature recognition, which is performed using advanced deep learning techniques, particularly Convolutional Neural Networks (CNNs). These CNNs are optimized for real-time processing on the GPU, significantly accelerating the analysis of facial features and the extraction of physiological signals such as heart rate and respiratory rate.

In this architecture, the GPU is not only responsible for rapid signal extraction but also for processing complex machine-learning algorithms, which are critical for the accurate estimation of physiological parameters. Parallel computing frameworks, such as OpenMP and OpenCL, are employed to further enhance system performance by enabling multi-threaded processing on both the CPU and GPU, ensuring efficient task scheduling and data handling. These frameworks allow for the dynamic allocation of resources, ensuring that bottlenecks in data transfer or processing are minimized, leading to smoother operation and faster results. Once the GPU completes the heavy computations, the CPU takes over for post-processing tasks, refining the extracted signals and performing any necessary filtering to enhance the accuracy of the physiological parameter estimations. The use of this heterogeneous architecture demonstrates a significant performance improvement over the homogeneous version, as it allows for faster processing time. The optimized system, therefore, offers a robust and efficient solution for real-time physiological monitoring, making it ideal for applications such as remote patient monitoring, telemedicine, and other non-invasive healthcare technologies where timely and accurate data processing is crucial (Table II).

TABLE II. HETEROGENEOUS SYSTEM IMPLEMENTATION

Function	Image acquisition and Preprocessing	Feature extraction	Parameter estimation
System	CPU	GPU	CPU/CPU
Sub-Function	Facial frames collecting	Signal extraction rPPG	Signal processing (GPU)
	Preprocessing	DL processing CNN	Post-processing (CPU)
Output	Preprocessed facial images	Extracted signals & features	Physiological parameter extraction

C. Performance Optimization

The performance of the proposed system was significantly enhanced through a combination of advanced optimization techniques aimed at improving both computational efficiency and real-time responsiveness. A key strategy involved leveraging parallel computing frameworks, with tasks distributed between the CPU and GPU to reduce bottlenecks and maximize resource utilization. OpenMP was employed to parallelize tasks on the CPU, enabling simultaneous execution of multiple processes, thereby reducing overall processing time. In parallel, OpenCL was utilized to harness the computational power of the GPU, particularly for tasks involving high-dimensional data processing, such as deep learning inference and signal extraction. This heterogeneous parallelism allowed the system to capitalize on the strengths of both processing units, with the CPU handling control and light processing tasks, while the GPU was responsible for more computationally intensive operations.

In addition to task distribution, memory management played a critical role in enhancing performance. To minimize data transfer overhead between the CPU and GPU, optimized memory allocation techniques were implemented, such as using pinned memory and efficient buffer management. This reduced latency associated with data movement and improved throughput. Furthermore, shared memory models were applied to accelerate data access and reduce cache misses during intensive computations.

Algorithmic optimizations were also a focus. For signal processing, advanced filtering techniques were used to accelerate the extraction of physiological parameters, while maintaining accuracy. In the deep learning component, the neural network models were optimized through pruning and quantization, reducing the model size and improving inference speed without compromising performance. These optimizations allowed the system to handle larger data inputs and deliver faster results, crucial for real-time monitoring applications.

To evaluate the impact of these optimizations, key performance metrics were assessed, including processing speed (in frames per second), accuracy of physiological parameter estimation, and resource utilization. The optimized heterogeneous system demonstrated significant improvements in processing speed compared to the naive homogeneous system, achieving real-time performance benchmarks. Resource utilization was carefully monitored to ensure efficient CPU-GPU collaboration, preventing bottlenecks and minimizing energy consumption. These optimizations not only enhanced the system's computational performance but also ensured scalability, making it suitable for deployment in real-time healthcare applications, such as remote patient monitoring and telemedicine.

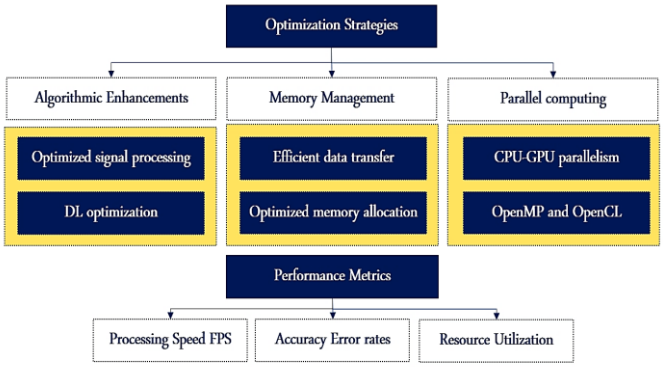


Fig. 10. Optimization strategies.

In the above Fig. 10, the optimization strategies for the system focus on enhancing performance through parallel computing, memory management, and algorithmic improvements. By distributing computational tasks between the CPU and GPU using frameworks like OpenMP and OpenCL, the system reduces bottlenecks and accelerates processing. Efficient data transfer and optimized memory allocation minimize overhead and speed up operations. Algorithmic enhancements include faster signal extraction methods and deep learning optimizations, such as model pruning and quantization, to improve inference speed. Performance metrics like processing speed (FPS), accuracy in physiological parameter estimation, and resource utilization are monitored to ensure effective and efficient operation.

V. Experimental Setup

A. Dataset and Experimental Protocol

The experimental dataset utilized in this study comprises RGB facial videos collected under various controlled and semi-controlled environmental conditions to accurately simulate real-world scenarios. These conditions include variations in ambient lighting, facial orientation, and subtle subject movements. Publicly available datasets such as UBFC-RPPG, COHFACE, or equivalent datasets were employed, each providing high-resolution facial videos paired with synchronized ground truth physiological signals, specifically heart rate and respiratory rate, obtained from reliable medical-grade sensors. Additionally, to enhance the robustness of the system and assess its performance under diverse conditions, a custom dataset was acquired using a high-definition camera (1080p resolution at 30 frames per second). In this setup, participants were positioned at a fixed distance of 1 to 2 meters from the camera, with uniform lighting to minimize external interferences.

The preprocessing phase involved converting the raw video sequences into individual frames, followed by face detection and tracking using advanced computer vision techniques, such as the Multi-task Cascaded Convolutional Networks (MTCNN) algorithm, to ensure precise extraction of the region of interest (ROI), specifically the facial area where physiological signals are most prominent. To ensure consistency across frames, facial landmarks were used to normalize the detected face, mitigating minor head movements and variations in facial orientation. For data alignment, the extracted video frames were synchronized with ground truth physiological signals through time-stamped data, ensuring accurate comparison during the validation phase. This synchronization allows for a one-to-one mapping between each frame and the corresponding physiological signal (e.g. heartbeats or respiratory cycles), which is crucial for training and testing the proposed system.

Once preprocessed, the dataset was split into training, validation, and testing subsets, ensuring a balanced distribution of conditions (e.g. lighting variations, subject movements) across all subsets. The preprocessed frames and synchronized ground truth signals were subsequently fed into the signal processing and deep learning pipelines. This comprehensive preprocessing and alignment ensured the system was rigorously tested across a variety of real-world scenarios, facilitating robust benchmarking of its performance in extracting physiological parameters such as heart rate and respiratory rate from facial images (Table III).

TABLE III. Dataset and Data Flow

Section	Subsection	Details
Dataset Overview	Dataset Composition	Facial images captured in varying conditions.
	Source	Open-source databases (e.g., UBFC-RPPG, COHFACE) or custom datasets with RGB videos.
	Custom Data Collection	- Frame rate (30 fps), resolution (1080p), distance (1-2 meters). - Description of lighting, subject movements, and position.
Experimental Protocol	Data Preprocessing	- Convert video sequences into frames for image processing. - Face detection (OpenCV, MTCNN) to extract facial regions, and normalize facial regions across frames for small movements.
	Ground Truth Alignment	- Synchronize physiological signals from wearable devices with video frames to establish ground truth.

B. Hardware and Software Setup

The experimental setup involves two distinct system configurations: a baseline homogeneous system and an optimized heterogeneous system, each designed to execute the same physiological parameter extraction tasks but under different architectural conditions. The homogeneous system, represents the naive implementation, which is typical of low-power embedded devices like the Raspberry Pi. In this configuration, all computational tasks, including image preprocessing, signal extraction, and deep learning-based prediction, are performed solely on the CPU without any hardware acceleration, thus providing a benchmark for performance evaluation. Conversely, the optimized system incorporates a heterogeneous architecture offering significant parallel processing capabilities. This architecture is employed to enhance computational efficiency by distributing workloads between the CPU and GPU. Specifically, OpenMP is used to parallelize tasks across multiple CPU cores, improving the efficiency of operations such as face detection and signal filtering, while OpenCL is utilized to offload computationally intensive tasks, such as deep learning inference and feature extraction, to the GPU. In both configurations, the software environment includes Python or C/C++ as the primary programming languages, alongside key libraries such as OpenCV for image and signal processing, TensorFlow/PyTorch for implementing deep learning models, and OpenMP/OpenCL to facilitate parallel processing. By combining these software tools with the respective hardware configurations, a comparative analysis of system performance, measured in terms of processing time, resource utilization, and overall computational efficiency, can be conducted, highlighting the advantages of heterogeneous architectures for real-time, non-contact physiological monitoring in embedded systems.

C. Signal Processing and Deep Learning Algorithm

The signal processing and deep learning algorithm for extracting physiological parameters from facial images involves a sophisticated multi-stage approach designed to enhance both accuracy and computational efficiency. The process begins with image preprocessing, where remote photoplethysmography (rPPG) techniques are employed to capture subtle, periodic color variations in the facial skin that correspond to physiological signals such as heart rate and respiration rate. This initial step involves extracting and aligning facial regions of interest from video frames using robust face detection algorithms. The extracted facial regions are then subjected to signal processing techniques to isolate the physiological signals from background noise. Specifically, a band-pass filter is applied to the raw photoplethysmographic signal to target the frequency bands associated with heart rate and respiration while filtering out high-frequency noise and low-frequency drift.

Following signal preprocessing, deep learning models are employed for advanced feature extraction and signal interpretation. Convolutional Neural Networks (CNNs) are utilized to analyze the spatial features of the facial images, enabling the system to recognize and extract features related to physiological changes. These features are then processed by Recurrent Neural Networks (RNNs), which are adept at handling time-series data and capturing temporal dependencies in the signal. The deep learning models are trained on a comprehensive dataset comprising facial images and corresponding ground truth physiological measurements, facilitating the learning of complex patterns and correlations between facial features and vital signs.

To optimize real-time performance, the system leverages heterogeneous computing architectures, integrating both CPU and GPU resources. Parallel processing frameworks, such as OpenMP, are employed to accelerate CPU-based tasks, including image preprocessing and feature extraction, while OpenCL is utilized to offload and expedite deep learning inference tasks to the GPU. This heterogeneous approach ensures efficient resource utilization and scalability, significantly reducing processing time compared to a homogeneous system. The combined use of advanced signal processing techniques and deep learning algorithms in a parallelized computing environment enables the system to achieve high accuracy in physiological parameter estimation while maintaining real-time operational capabilities, making it highly suitable for applications in remote health monitoring and telemedicine.

D. Evaluation Metrics

In evaluating the performance of the physiological parameter extraction systems, we employ a multifaceted approach that encompasses both accuracy and efficiency aspects. Accuracy Metrics focus on quantifying the precision of physiological parameters extracted from facial images. Key metrics include:

1) Mean Absolute Error (MAE): This metric measures the average magnitude of errors between the extracted and ground truth physiological parameters. It provides a straightforward indication of the system's accuracy in estimating parameters such as heart rate and respiratory rate.

2) Root Mean Square Error (RMSE): RMSE evaluates the square root of the average squared differences between extracted values and ground truth. This metric is particularly useful for assessing the impact of larger deviations and provides insight into the consistency and reliability of the parameter estimations.

Performance Metrics assess the operational efficiency and speed of the systems:

3) Processing time: This metric measures the elapsed time required for the system to process a sequence of images or video frames. It is critical for evaluating the system's capability to operate in real-time, with lower processing times indicating enhanced performance.

4) Resource utilization: This involves monitoring CPU and GPU usage during system operation. Efficient resource utilization is essential for optimizing system performance, particularly in heterogeneous systems where balancing computational load between CPU and GPU can significantly impact overall efficiency.

By analyzing these metrics, we gain comprehensive insights into both the accuracy of physiological parameter extraction and the operational efficiency of the systems. This evaluation not only highlights the strengths and limitations of the homogeneous and heterogeneous implementations but also provides a detailed understanding of their practical applicability in real-world scenarios. The comparative analysis informs decisions on optimizing system design for enhanced performance and reliability in non-invasive health monitoring applications. The following Tables IV to VI illustrate the eventual evaluation metrics overview done, and the values outlined by the MAE and the RMSE.

TABLE IV. EVALUATION METRICS OVERVIEW

Metric	Homogeneous System	Heterogeneous System
Processing Time	120ms per frame	45ms per frame
Accuracy (HR)	± from 1 to 2 bpm	
Accuracy (RR)	±from 1 to 3 breaths per minute	
Resource Utilization	High CPU usage	Balanced CPU/GPU usage

TABLE V. MEAN ABSOLUTE ERROR (MAE)

Parameter	Homogeneous System	Heterogeneous System
Heart Rate	~From 1.8 to 3.5	
Respiratory Rate	~From 0.9 to 2.1	

TABLE VI. ROOT MEAN SQUARE ERROR (RMSE)

Parameter	Homogeneous System	Heterogeneous System
Heart Rate	~From 2.0 to 4.2	
Respiratory Rate	~From 1.1 to 2.5	

VI. Results and Analysis

A. Experimental Setup and Dataset

The experimental setup for evaluating the proposed physiological parameter extraction system is meticulously designed to assess performance across different configurations. The dataset comprises a diverse set of facial images, encompassing a wide range of lighting conditions, facial expressions, and orientations. These images a captured in controlled environments to ensure variability and robustness. Each image within the dataset is meticulously annotated with ground truth values for physiological parameters such as heart rate and respiratory rate, enabling precise validation of the system's accuracy.

The experimental environment includes two distinct computing platforms: a naive homogeneous system and an optimized heterogeneous system. The homogeneous system operates on a single type of processor, serving as the baseline for performance comparison. In contrast, the heterogeneous system leverages a combination of CPU and GPU resources, utilizing parallel processing frameworks such as OpenMP and OpenCL to enhance computational efficiency and real-time processing capabilities. Detailed specifications of both systems are documented, including processor models, memory configurations, and software environments.

The experimental workflow encompasses several stages: image acquisition, preprocessing, feature extraction, and parameter estimation. Preprocessing steps involve image normalization, noise reduction, and enhancement to standardize input data. Feature extraction employs advanced signal processing techniques to isolate relevant facial features used for physiological parameter estimation. Deep learning algorithms, including convolutional neural networks (CNNs), are then applied to extract and predict the desired parameters. Diagrams illustrating the experimental setup, such as camera positioning and system architecture, are provided to represent the setup.

B. Performance Metrics

Processing time metrics are critical for understanding the system's capability to perform in real-time scenarios. Metrics such as the average processing time per image reflect the time required to process a single facial image and extract the necessary physiological parameters. The total processing time for a batch of images is also assessed to evaluate the system's efficiency in handling multiple inputs simultaneously. These metrics help in identifying potential delays and ensuring that the system can meet the real-time requirements of practical applications.

Resource utilization metrics provide insight into the efficiency with which the system uses computational resources. CPU utilization measures the percentage of processing power utilized by the central processing unit, while GPU utilization assesses the usage of the graphics processing unit, crucial for systems leveraging heterogeneous architectures. Memory usage metrics track the amount of RAM consumed during processing, which can influence the system's ability to handle large datasets or perform complex computations. Analyzing these metrics allows for the identification of resource bottlenecks and opportunities for optimization, ensuring that the system operates efficiently within the constraints of the hardware.

Together, these performance metrics offer a comprehensive evaluation of the system's ability to accurately and efficiently extract physiological parameters, highlighting areas for improvement and optimization. By systematically analyzing these metrics, the study provides a clear picture of the system's strengths and limitations, facilitating informed decisions on further enhancements and practical deployment in healthcare applications.

C. Comparative Analysis

The comparative analysis systematically evaluates the performance disparities between the naive homogeneous system and the optimized heterogeneous system, focusing on processing efficiency, accuracy, and resource utilization. In terms of processing efficiency, the optimized heterogeneous system, which utilizes a CPU-GPU architecture, demonstrates a significant reduction in image processing time. This improvement is largely attributable to the parallel processing capabilities enabled by frameworks such as OpenMP and OpenCL, which facilitate concurrent execution of computational tasks and efficient utilization of available hardware resources. The analysis also considers resource utilization, where the heterogeneous system exhibits superior efficiency in managing computational resources. CPU and GPU utilization metrics indicate that the optimized system achieves higher throughput and lower idle times.

Overall, the results highlight the tangible benefits of adopting a heterogeneous architecture for real-time physiological parameter extraction. The optimized system not only accelerates processing but also maintains higher accuracy, making it a more effective solution for demanding healthcare monitoring applications. This comparative analysis underscores the importance of leveraging advanced parallel computing techniques to achieve significant performance gains in complex real-time systems. In the following Fig. 11 illustrates the processing time achieved using the naïve version and the optimized version versus the number of frames. Also, Fig. 12 and 13 show the estimated heart and respiratory rates versus the actual ones.

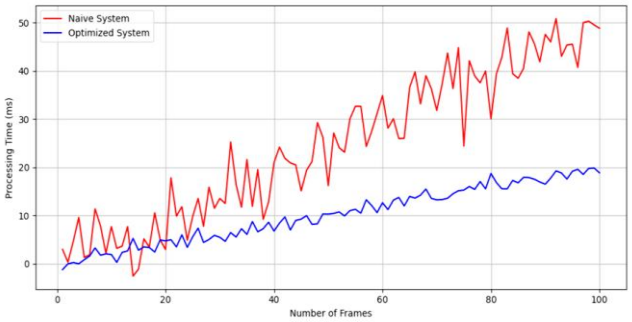


Fig. 11. Processing time vs. Number of frames.

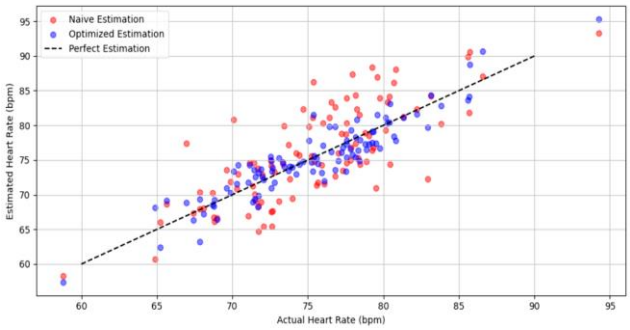


Fig. 12. Estimated HR vs. Actual HR.

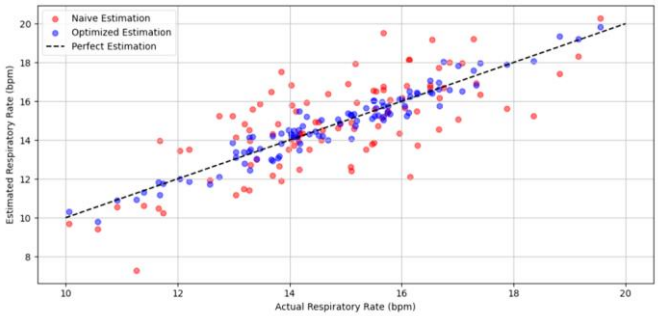


Fig. 13. Estimated RR vs. Actual RR.

VII. Conclusion

This study presents a groundbreaking approach for the non-contact extraction of physiological parameters, such as heart rate and respiratory rate, from facial images using RGB cameras, capitalizing on advanced deep learning and signal processing methodologies. The proposed system adeptly addresses several key challenges, including variations in lighting conditions, facial orientation, and background noise, through the integration of sophisticated AI-driven algorithms. A comprehensive evaluation reveals that the optimized heterogeneous architecture, employing both CPU and GPU resources, significantly outperforms the traditional homogeneous system in terms of processing speed and computational efficiency. The optimization achieved through parallel computing frameworks, notably OpenMP and OpenCL, results in marked improvements in real-time performance while preserving high accuracy levels. This advancement underscores the efficacy of heterogeneous architectures in enhancing the scalability and responsiveness of non-invasive physiological monitoring systems. The study's findings not only validate the potential of these technologies for applications in remote patient monitoring and telemedicine but also highlight the importance of continued innovation in system design and computational techniques to meet the evolving needs of healthcare technology. The successful implementation and demonstrated performance of the proposed system represent a significant step forward in the field of remote health monitoring, providing a robust platform for future research and development.

VIII. Future Work

Future work should focus on several critical aspects to advance the capabilities and practical application of non-contact physiological monitoring systems. Firstly, there is a need for continued refinement of the deep learning models and signal processing algorithms employed in the system. This includes exploring techniques such as model pruning and quantization to enhance computational efficiency and reduce latency without compromising accuracy. Additionally, integrating multi-modal sensing technologies, such as combining facial image analysis with thermal imaging or data from wearable sensors, could significantly improve the robustness and precision of physiological measurements by providing complementary data that address limitations inherent in single-modal systems. Extensive field testing across varied environments and diverse demographic groups is also essential to validate the system's performance in real-world conditions, ensuring reliability and adaptability under different lighting conditions, facial orientations, and levels of noise. Furthermore, optimizing the system for deployment on mobile and edge computing platforms would increase its accessibility and usability, making it more practical for widespread adoption. Lastly, addressing ethical and privacy concerns related to the collection and use of facial data is of utmost importance. This involves developing comprehensive guidelines and implementing advanced technologies to protect user data, ensure informed consent, and uphold privacy standards. By tackling these areas, future research can build upon the current advancements, pushing the boundaries of non-contact physiological monitoring and contributing to more effective, efficient, and ethical healthcare solutions.

Acknowledgment

We owe a debt of gratitude to the Ministry of National Education, Vocational Training, Higher Education and Scientific Research (MENFPESRS) and the National Centre for Scientific and Technical Research of Morocco (CNRST) for their financial support for the project Cov/2020/109.

References

[1] Molinaro, N., Schena, E., Silvestri, S., Bonotti, F., Aguzzi, D., Viola, E., . & Massaroni, C. (2022). Contactless vital signs monitoring from videos recorded with digital cameras: an overview. Frontiers in Physiology, 13, 801709.

[2] Liu, X., Patel, S., & McDuff, D. (2021). RGB Camera-based Physiological Sensing: Challenges and Future Directions. arXiv preprint arXiv:2110.13362.

[3] Rouast, P. V., Adam, M. T., Chiong, R., Cornforth, D., & Lux, E. (2018). Remote heart rate measurement using low-cost RGB face video: a technical literature review. Frontiers of Computer Science, 12, 858-872.

[4] Okada, G., Yonezawa, T., Kurita, K., & Tsumura, N. (2018). Monitoring emotion by remote measurement of physiological signals using an RGB camera. ITE Transactions on Media Technology and Applications, 6(1), 131-137.

[5] Selvaraju, V., Spicher, N., Wang, J., Ganapathy, N., Warnecke, J. M., Leonhardt, S., ... & Deserno, T. M. (2022). Continuous monitoring of vital signs using cameras: A systematic review. Sensors, 22(11), 4097.

[6] Przybyło, J. (2022). A deep learning approach for remote heart rate estimation. Biomedical Signal Processing and Control, 74, 103457.

[7] Lee, R. J., Sivakumar, S., & Lim, K. H. (2024). Review on remote heart rate measurements using photoplethysmography. Multimedia Tools and Applications, 83(15), 44699-44728.E

[8] Przybyło, J. (2022). A deep learning approach for remote heart rate estimation. Biomedical Signal Processing and Control, 74, 103457.

[9] Liu, F., Chen, D., Zhou, J., & Xu, F. (2022). A review of driver fatigue detection and its advances on the use of RGB-D camera and deep learning. Engineering Applications of Artificial Intelligence, 116, 105399.

[10] Yu, J., de Antonio, A., & Villalba-Mora, E. (2022). Deep learning (CNN, RNN) applications for smart homes: a systematic review. Computers, 11(2), 26.

[11] Ghanadian, H., Ghodratigohar, M., & Al Osman, H. (2018). A machine learning method to improve non-contact heart rate monitoring using an RGB camera. IEEE Access, 6, 57085-57094.

[12] Gupta, A., Ravelo-García, A. G., & Morgado-Dias, F. (2024). Recent advancements in deep learning-based remote photoplethysmography methods. Data Fusion Techniques and Applications for Smart Healthcare, 127-155.

[13] Nocera, A., Senigagliesi, L., Raimondi, M., Ciattaglia, G., & Gambi, E. (2021). Machine Learning in RADAR-based Physiological Signals Sensing: A Scoping Review of the Models, Datasets and Metrics. Machine Learning, 19, 1.

[14] Cheng, F. H., & Chen, Y. L. (2006). Real time multiple objects tracking and identification based on discrete wavelet transform. Pattern recognition, 39(6), 1126-1139.

[15] Al-Berry, M. N., Salem, M. M., Hussein, A. S., & Tolba, M. F. (2013, November). Motion detection using wavelet-enhanced accumulative frame differencing. In 2013 8th International Conference on Computer Engineering & Systems (ICCES) (pp. 255-261). IEEE.

[16] Anwer, S., Li, H., Antwi-Afari, M. F., Mirza, A. M., Rahman, M. A., Mehmood, I., ... & Wong, A. Y. L. (2024). Evaluation of data processing and artifact removal approaches used for physiological signals captured using wearable sensing devices during construction tasks. Journal of Construction Engineering and Management, 150(1), 03123008.

[17] Mittal, S., & Vetter, J. S. (2015). A survey of CPU-GPU heterogeneous computing techniques. ACM Computing Surveys (CSUR), 47(4), 1-35.

[18] Rakhimov, M., & Ochilov, M. (2021, October). Distribution of operations in heterogeneous computing systems for processing speech signals. In 2021 IEEE 15th International Conference on Application of Information and Communication Technologies (AICT) (pp. 1-4). IEEE.

[19] Tu, Y., Sadiq, S., Tao, Y., Shyu, M. L., & Chen, S. C. (2019, July). A power efficient neural network implementation on heterogeneous FPGA and GPU devices. In 2019 IEEE 20th international conference on information reuse and integration for data science (IRI) (pp. 193-199). IEEE.

[20] Rodriguez, D., Gomez, D., Alvarez, D., & Rivera, S. (2021). A review of parallel heterogeneous computing algorithms in power systems. Algorithms, 14(10), 275.

[21] El Boussaki, H., Latif, R., Saddik, A. (2023, April). A Review on Video-Based Heart Rate, Respiratory Rate and Blood Pressure Estimation. In Advances in Machine Intelligence and Computer Science Applications: Proceedings of the International Conference ICMICSA'2022 (pp. 129-140). Cham: Springer Nature Switzerland.

[22] El Khadiri, Z., Latif, R., Saddik, A. (2023, March). Breathing Pattern Assessment Through the Empirical Mode Decomposition and the Empirical Wavelet Transform Algorithms. In The 3rd International Conference on Artificial Intelligence and Computer Vision (AICV2023), March 5–7, 2023 (pp. 262-271). Cham: Springer Nature Switzerland.

[23] Latif, R., Addaali, B., Saddik, A. (2023). Real-Time SPO2 Monitoring Based on Facial Images Sequences. In Digital Technologies and Applications: Proceedings of ICDTA'23, Fez, Morocco, Volume 1 (pp. 474-483). Cham: Springer Nature Switzerland.

[24] El Boussaki, H., Latif, R., Saddik, A., El Khadiri, Z., El Boujaoui, H. (2023). Non-contact Respiratory Rate Monitoring Based on the Principal Component Analysis. International Journal of Advanced Computer Science and Applications, 14(9).

[25] El khadiri, Z., Latif, R., Saddik, A. (2022, November). Remote Heart Rate Measurement Using Plethysmographic Wave Analysis. In International Conference of Machine Learning and Computer Science Applications (pp. 254-267). Cham: Springer Nature Switzerland.

[26] El Khadiri, Z., Latif, R., & Saddik, A. (2023). An efficient hybrid algorithm for non-contact physiological sign monitoring using plethysmography wave analysis. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 11(6), 2305-2321.

[27] Cha, Y. J., Ali, R., Lewis, J., & Büyüköztürk, O. (2024). Deep learning-based structural health monitoring. Automation in Construction, 161, 105328.

[28] Sharma, S. K., & Dixit, R. J. (2024). Applications of Parallel Data Processing for Biomedical Imaging. In Applications of Parallel Data Processing for Biomedical Imaging (pp. 1-24). IGI Global.

[29] EL BOUSSAKI, H. O. D. A., LATIF, R., & SADDIK, A. (2024). REMOTE MONITORING OF THE DRIVER'S HEART RATE, BREATHING RATE AND FATIGUE SIGNS USING A HYBRID APPROACH ON DIFFERENT EMBEDDED ARCHITECTURES. Journal of Theoretical and Applied Information Technology, 102(18).

[30] Hosni, A., & Atef, M. (2023). Remote real-time heart rate monitoring with recursive motion artifact removal using PPG signals from a smartphone camera. Multimedia Tools and Applications, 82(13), 20571-20588.

[31] Di Lernia, D., Finotti, G., Tsakiris, M., Riva, G., & Naber, M. (2024). Remote photoplethysmography (rPPG) in the wild: Remote heart rate imaging via online webcams. Behavior Research Methods, 1-11.

[32] Xiao, H., Liu, T., Sun, Y., Li, Y., Zhao, S., & Avolio, A. (2024). Remote photoplethysmography for heart rate measurement: A review. Biomedical Signal Processing and Control, 88, 105608.

[33] Karthick, R., Dawood, M. S., & Meenalochini, P. (2023). Analysis of vital signs using remote photoplethysmography (RPPG). Journal of Ambient Intelligence and Humanized Computing, 14(12), 16729-16736.

[34] Rai, M., Yadav, R. K., Husain, A. A., Maity, T., & Yadav, D. K. (2018). Extraction of Facial Features for Detection of Human Emotions under Noisy Condition. International Journal of Engineering and Manufacturing, 8(5), 49.

[35] Jerritta, S., Murugappan, M., Nagarajan, R., & Wan, K. (2011, March). Physiological signals based human emotion recognition: a review. In 2011 IEEE 7th international colloquium on signal processing and its applications (pp. 410-415). IEEE.

[36] Egger, M., Ley, M., & Hanke, S. (2019). Emotion recognition from physiological signal analysis: A review. Electronic Notes in Theoretical Computer Science, 343, 35-55.

[37] Vuorela, T., Seppä, V. P., Vanhala, J., & Hyttinen, J. (2010). Design and implementation of a portable long-term physiological signal recorder. IEEE transactions on information technology in biomedicine, 14(3), 718-725.

[38] Lin, K. C., & Fang, W. C. (2014, May). A highly integrated hardware design implemented on FPGA for a wireless healthcare monitoring system. In 2014 IEEE International Conference on Consumer Electronics-Taiwan (pp. 187-188). IEEE.

[39] Dilmaghani, R. S., Bobarshad, H., Ghavami, M., Choobkar, S., & Wolfe, C. (2011). Wireless sensor networks for monitoring physiological signals of multiple patients. IEEE Transactions on biomedical circuits and systems, 5(4), 347-356.

[40] Da Silva, H. C., Pisani, F., & Borin, E. (2016, October). A comparative study of SYCL, OpenCL, and OpenMP. In 2016 International Symposium on Computer Architecture and High Performance Computing Workshops (SBAC-PADW) (pp. 61-66). IEEE.

[41] Shen, J., Fang, J., Sips, H., & Varbanescu, A. L. (2012, September). Performance gaps between OpenMP and OpenCL for multi-core CPUs. In 2012 41st International Conference on Parallel Processing Workshops (pp. 116-125). IEEE.

[42] Thouti, K., & Sathe, S. R. (2012). Comparison of OpenMP & OpenCL parallel processing technologies. arXiv preprint arXiv:1211.2038.

Smart Muni Platform: Efficient Emergency and Citizen Security Management Based on Geolocation, Technological Integration and Real Time Communication

Gerson Castro-Chucan, Hiroshi Chalco-Peñafiel, Norka Bedregal-Alpaca, Victor Cornejo-Aparicio
Universidad Nacional de San Agustín de Arequipa, Arequipa, Perú

***Abstract*—The global increase in crime across cities has led to the development and implementation of technological solutions, such as the Smart Muni platform, designed to enhance citizen security. This platform integrates geolocation, real-time notifications, and a digital panic button to optimize emergency management and coordination between citizens and authorities. Developed using a structured approach that included requirements analysis, system design, development, testing, validation, deployment, and maintenance, the platform employs advanced technologies such as Firebase, Amazon S3, and Twilio, ensuring scalability, high availability, and seamless communication. Initially implemented in two districts with high crime rates, Smart Muni registered an average of 10 daily alerts, with peaks of up to 50 alerts in a single day. The system has proven effective in managing frequent incidents like alcoholism and domestic violence, significantly reducing response times and improving coordination. Despite its success, Smart Muni faces challenges related to optimizing its resilience against potential system failures and improving its ability to handle increased data loads. In comparison to other international systems, Smart Muni's flexible and scalable architecture stands out, though future enhancements are required to further strengthen the system's reliability and expand its features. Overall, Smart Muni has proven to be a valuable tool in improving citizen security, fostering stronger relationships between citizens and authorities, and contributing to a safer community.**

***Keywords*—Citizen security; geolocation; emergency management; cloud technology; alert platform**

I. Introduction

Citizen security refers to the protection of citizens' rights and freedoms, as well as the prevention and control of violence and crime in urban environment [1]. For García and Zambrano [2], this concept should be conceived from a broad perspective, involving both the perception of security by the population and the concrete actions carried out by the authorities to guarantee it.

Globally, many cities are experiencing rising crime rates and facing major challenges in emergency management. Rapid urbanization and increased population density in urban areas have aggravated these problems, highlighting the need for innovative solutions to ensure public safety. In large cities, the adoption of advanced technologies has proven to be effective in improving coordination and response to crime incidents and emergencies; to increase the effectiveness and efficiency of public safety operations, innovative strategies using the Global Positioning System (GPS) have been implemented. Cities such as New York and London have integrated geolocation systems and data analysis into their security strategies, allowing for a faster and more accurate response to emergencies. However, in Latin American countries, the integration of technology to promote the security of their citizens is lagging behind [3].

While several international systems, such as those implemented in New York and London, have successfully integrated geolocation and data analysis to improve emergency response, these solutions often focus on larger, well-resourced urban centers. In contrast, Latin American cities, particularly in Peru, face unique challenges such as fragmented communication infrastructures and limited access to technological resources. The existing systems also lack the ability to address the specific needs of local municipalities, which often operate under constrained budgets and require more adaptable solutions. Smart Muni fills this gap by providing a scalable, cost-effective platform tailored to the context of smaller cities in developing regions. Its integration of real-time communication, geolocation, and cloud-based technologies ensures that it can quickly adapt to the evolving needs of these communities, while also offering enhanced resilience in regions where emergency response systems are often under-resourced.

The problem of citizen security in Peru is complex and multifaceted; many Peruvian cities are experiencing an increase in crime rates, a situation that is affected by various social, economic and political factors. According to the National Strategic Programs Survey, nine out of 10 people believed that they would be victims of crime in the next 12 months [4]. Citizens and local authorities face difficulties in coordinating quick and effective actions to guarantee citizen security; there is a growing concern about acting in a timely manner to prevent crime or mitigate its consequences.

The lack of an integrated incident management system and the absence of adequate technological tools for the reception and management of security and emergency alerts compromises the authorities' ability to act in a timely manner and reduces community confidence in public safety, demonstrating the urgent need to implement an advanced technological solution to optimize emergency response and management, improve communication between citizens and authorities, and strengthen

citizen security in the region. Sarkar [5] and Samarakkody et al. [6] suggest that technological innovations enhance disaster resilience and empower citizens, creating livable communities and improving overall quality of life. Protecting a complex environment, such as a city, is a task of enormous and increasing complexity, requiring the combination of multidisciplinary techniques and tools, the availability of very diverse skills, and the ability to correlate and explore large flows of data and information [7]. Technology plays a crucial role in smart cities, supporting the optimization of urban services and improving the quality of life of their inhabitants [8].

Several research studies explore the application of geolocation and geographic information systems (GIS) to improve citizen security and tourism safety. Alvarez [9] proposes extending Ecuador's ECU 9-1-1 system to rural areas, incorporating georeferencing and monitoring cameras to enhance emergency response. In Mexico, Espinoza-Ramirez et al. [10] developed a mobile device-based system using GIS to map crime incidents, helping citizens to identify and avoid dangerous areas. However, Gonzalez [11] raises concerns about the constitutionality of citizen geolocation through mobile applications, particularly about the rights to privacy and freedom of movement. Solano Barliza [12] suggests that combining recommender systems with geolocation technologies could improve tourism safety, pointing out the lack of existing solutions specifically focused on this aspect.

The implementation of a geolocation-based software application to detect aid points when a citizen observes or suffers a crime is crucial for a rapid and coordinated response by the authorities [13], which would significantly improve communication between citizens and authorities, strengthen trust in public safety, and empower the community to actively participate in their own protection.

For their part, Jesus et al. [14] deal with the reliability and detectability of emergency management systems, emphasizing the importance of resilience to common failures. Costa et al. [15] present a review of various emergency management systems, highlighting the diversity of technologies and approaches used. Samarakkody et al. [16] investigate technological innovations that improve disaster resilience in smart cities.

In this information context, a robust and efficient platform for the management of citizen emergencies and alerts is proposed, which will significantly improve the response and coordination capacity between citizens and authorities and allow for a more direct and real-time interaction, facilitating decision making and the mobilization of resources to efficiently respond to citizens' calls for help.

II. Methodology

A. Stages

In order to develop a robust and effective platform to improve citizen security through the use of geolocated technology, a structured methodology has been followed that includes the following stages:

1) Requirements analysis: Identification of needs and definition of functionalities.

2) System design: Definition of the technical structure and design of an intuitive and accessible interface for users.

3) Development: Identification of suitable languages and frameworks for web and mobile development.

4) Testing and validation: Performing unit and integration testing to ensure that each component works correctly and the app as a whole operates smoothly, and user testing to get feedback and make necessary adjustments.

5) Deployment and maintenance: Release, monitoring and updates to improve functionality and security.

B. Implementation Context

The proposed platform was implemented in the districts of Cayma and José Luis Bustamante y Rivero, located in the region of Arequipa, Peru; districts that face a significant increase in crime. With populations of approximately 91,935 and 82,642 inhabitants, respectively, these districts have seen an increase in crimes such as robberies, vandalism, family violence and cases of alcoholism. In 2022, more than 1290 robberies and about 1038 cases of family violence were reported [17][18].

The growing concern among citizens and authorities is due to the inadequate and delayed response to emergencies, aggravated by the absence of a unified incident management system. To deal with this problem, Smart Muni was implemented, a comprehensive platform consisting of a web application, a mobile application for the sereno (municipal security officer) and a mobile application for the citizen. This solution seeks to improve coordination and efficiency in emergency attention, facilitating communication and incident management in the districts of Cayma and José Luis Bustamante y Rivero.

III. Platform Description

The direct users of the platform are citizens, serenos and municipalities (Fig. 1). Citizens use a mobile application to report incidents and emergencies, providing details such as location and multimedia information. Additionally, they receive notifications about their reports, access security information, use the panic button in emergency situations and consult the incident history.

For their part, the serenos, members of the citizen security corps, work to maintain order and security in streets and public spaces. Employed by local governments, such as provincial and district municipalities, their main function is to prevent and control risk situations, such as robberies, aggressions or disturbances. The serenos, using the corresponding mobile application, manage the incidents reported by citizens through geolocated alerts, resolving and updating them in real time, and they are monitored to ensure their location.

The municipalities use a web application to manage the platform. This web application allows attending to alerts or incidents and assigning them to the corresponding units, analyzing data in real time through advanced analysis tools to identify patterns and improve security, and sending notifications to citizens in a timely manner. Furthermore, municipalities can manage user permissions and access, integrate the system with other government and emergency entities, and continuously

monitor all activities to ensure a quick and efficient response to any incident.

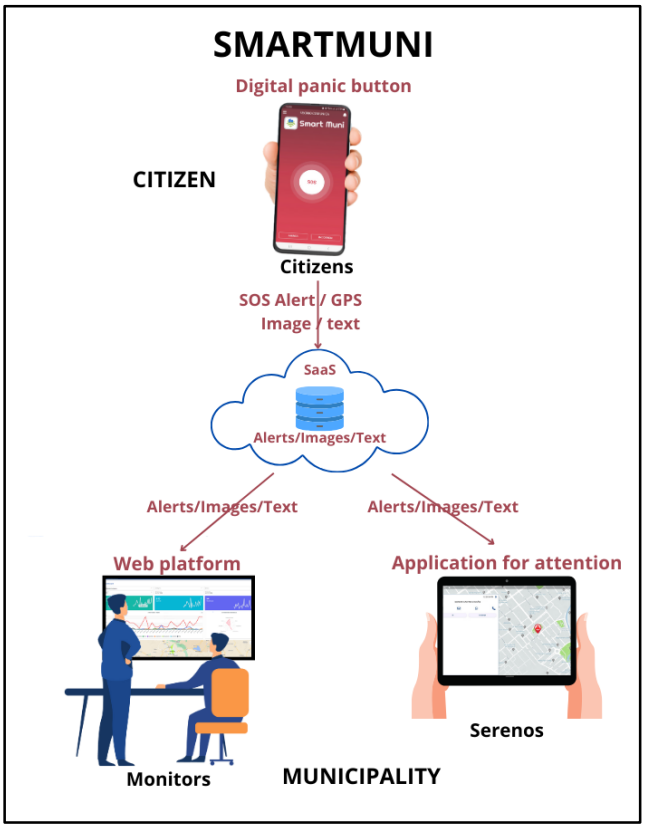


Fig. 1. Smart muni scheme.

A. Integration of Technologies

The platform architecture was designed and developed internally to ensure efficient control over the deployment and updating of the project. Fig. 2 shows the technologies employed in three key areas: data storage, message sending and deployment.

1) Data storage: For data storage, Amazon S3, PostgreSQL and Firebase Realtime Database technologies were used [19][20][21].

2) Deployment: Docker, Docker Compose and GitLab CI/CD along with GitLab Runner were used for developing, sending and executing containerized applications, facilitating continuous integration and delivery [22][23][24][25].

3) Web service providers: Web services were provided by DigitalOcean and GoDaddy, simplifying IT infrastructure and offering online tools for business [26][27].

4) Messaging: Firebase Cloud Messaging and Twilio were used for messaging and notifications, enabling communications on iOS, Android devices and web applications [28][29].

5) Development technologies: Django REST Framework, Quasar Framework and Flutter were used in the development of technologies, allowing the construction of web APIs, user interfaces and hybrid mobile applications [30][31][32].

6) Servers and middleware: Servers and middleware included Nginx, uWSGI, Let's Encrypt and Ubuntu UFW, providing high performance, security and firewall rule management on servers [33][34][35][36].

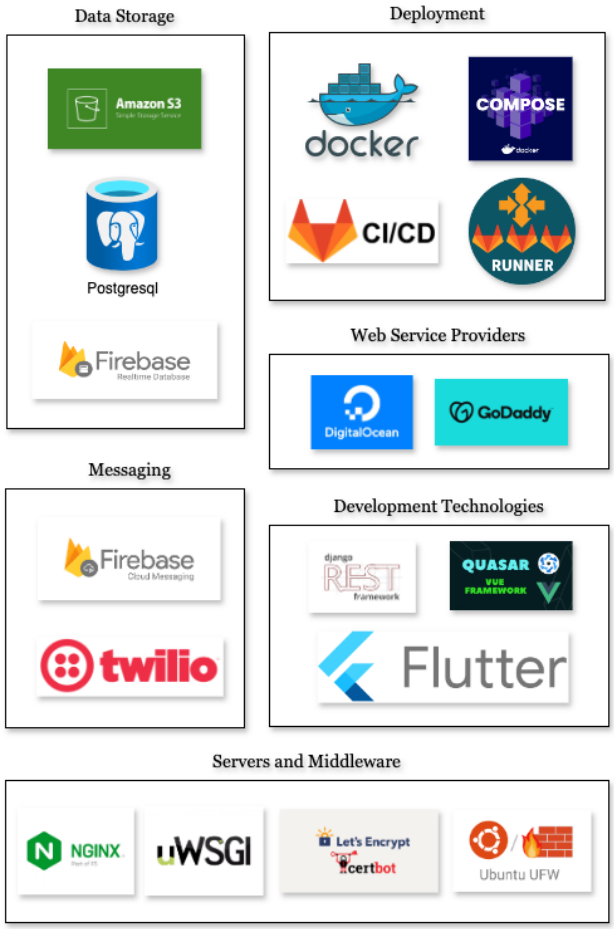


Fig. 2. Scheme of the used technologies.

B. Software Architecture

Fig. 3 shows the software architecture of the platform, integrating several external services and specialized databases.

The platform architecture is composed of the following elements:

1) Firebase services: Includes Firebase Real Time Database and Firebase Cloud Messaging for real-time storage and sending notifications to mobile devices.

2) Google service: Google Maps APIs for mapping and location functionalities.

3) Amazon service: Amazon S3 for file and data storage.

4) Twilio service: For communication and sending SMS messages.

5) GeoLite2 service: For geographic location of IP addresses.

6) PostgreSQL database: Used as the main data store.

7) Web services: This central component manages the interactions between models, controllers, views and the ORM. It is the system core that connects database and external services with the mobile applications and the website.

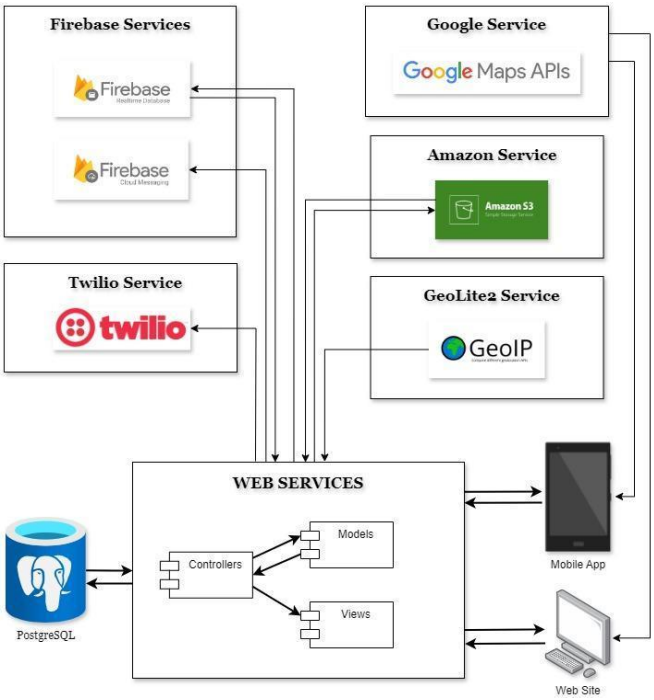


Fig. 3. Platform architecture.

This architecture allows the integration of several services and technologies providing a robust and scalable solution that supports both mobile and web applications. The web application uses Vue.js (progressive framework for building user interfaces) for client development, structured with VueRouter (library for handling navigation in Vue.js applications) for navigation and a Store (centralized solution for managing the state of an application) for state management (Fig. 4).

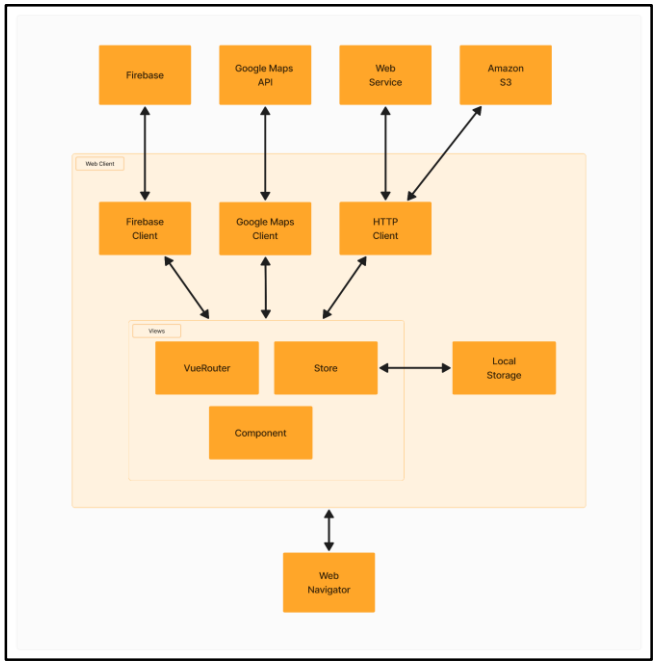


Fig. 4. Web architecture.

These components interact with Firebase for real-time databases, Google-Maps-API for maps, an API for data and Amazon S3 for file storage through specialized clients. The browser's local storage is used for session data. User interaction is performed through a web browser, communicating directly with Vue.js components for an interactive and dynamic experience.

For the mobile architecture (Fig. 5), Flutter (UI framework for creating native applications) is used with the BLoC design pattern to manage events and states. Flutter components communicate with Firebase for real-time databases, Google Maps API for maps and a web service to consume data from an API through specialized clients. The SharedPreferences library is used to persist data locally, and mobile services such as GPS and camera are integrated. The application, running on Android and iOS devices, interacts with Flutter widgets, efficiently managing resources and functionality.

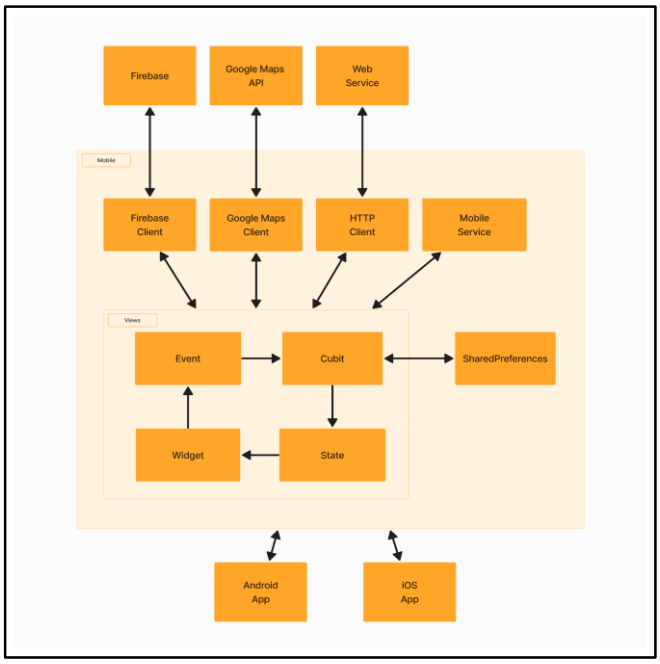


Fig. 5. Mobile architecture.

C. Navigability Diagrams

The web application allows the comprehensive management of various modules related to citizen security, including incidents, alerts, monitoring and events. Moreover, it facilitates the administration of venues, advertising, geozones and other municipal aspects, both on the part of the municipality and the administrators. The assignment of specific permissions according to account type is also considered (Fig. 6).

The mobile application for citizens (Fig. 7) allows the creation of accounts, sending alerts and generating incidents through an interactive map. Besides, configuration of profiles and the possibility of reviewing the history of alerts and incidents are also available.

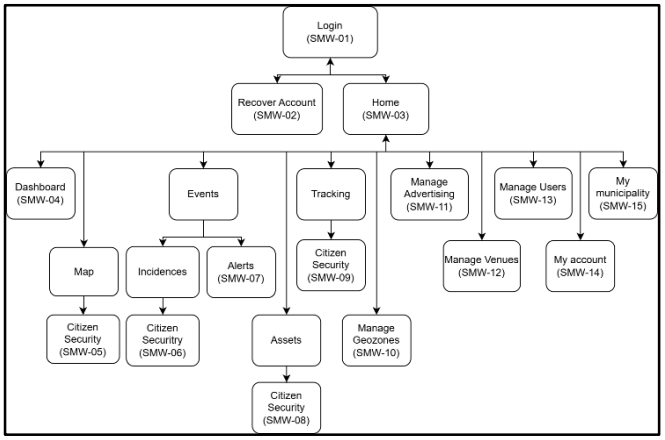


Fig. 6. Web application navigability diagram.

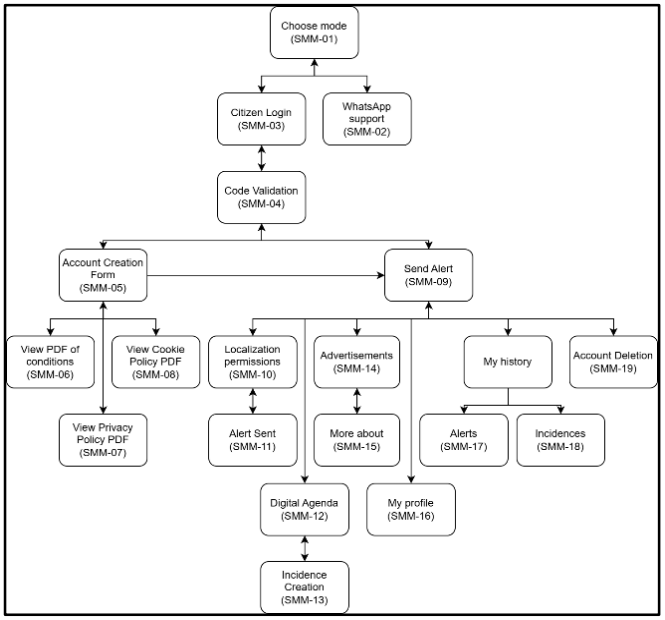


Fig. 7. Citizen's application navigability diagram.

The mobile application for serenos (Fig. 8) allows the configuration of profiles and locations, as well as the review of the history of alerts and incidents of all users. It also facilitates visualization of this information through PDF files.

D. Interfaces

The user interface is the point of human-computer interaction and communication on a device. Fig. 9 and Fig. 10 show main interfaces of the platform. Fig. 9 shows the main screen of Smart Muni web application, where an interactive map showing locations of different incidents and alerts in the city can be seen.

In this interface, alerts and incidents can be attended, as well as incidents can be filtered by their status (attended or not attended) and by type of incident (alert, family violence, alcoholism, drug addiction, robbery, vandalism, accident and others). In addition, monitoring by application is displayed, showing the number of workers, motorcycles, vans and other vehicles monitored in real time, as well as the number of cameras and sirens available.

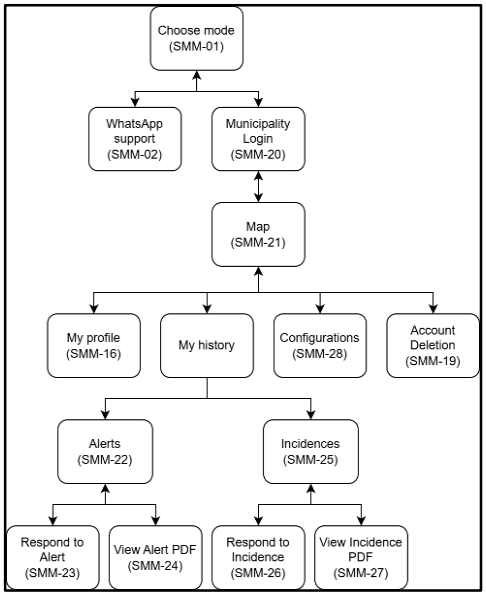


Fig. 8. Sereno's application navigability diagram.

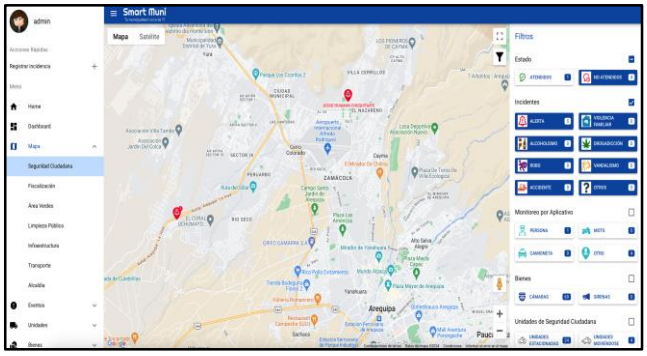


Fig. 9. Main interface of the control module and display of alerts and incidents.

Fig. 10. Mobile app sending an alert.

Detailed interfaces can be accessed by clicking on: INTERFACES.pdf

IV. RESULTS

The implementation of Smart Muni in the districts of Cayma and José Luis Bustamante y Rivero has significantly improved citizen security, demonstrating its positive impact on management of emergencies and citizen reports.

Smart Muni in the district of José Luis Bustamante has been in operation since December 20, 2023, and in the district of Cayma since June 20, 2023.

The historical record made by the application is fundamental to evaluate the accuracy and efficiency of the system compared to the manual methods previously used, methods that, since they are not adequately regulated, have not allowed access to the historical data recorded manually. Data collection through the Smart Muni platform allows for the creation of a solid base of information to be analyzed over time. This continuous recording ensures data integrity, and also facilitates identification of patterns, trends and potential improvements in the process. The historical data collected by the application over one year (Table I) are evidence of the system's impact and evolution, emphasizing its capacity to generate accurate and consistent records that contribute to informed decision making.

TABLE I. NUMBER OF ALERTS AND INCIDENCES PER MONTH

Month / Year	Cayma	José Luis Bustamante y Rivero
August 2023	145	-
September 2023	128	-
October 2023	89	-
November 2023	93	-
December 2023	117	156
January 2024	56	99
February 2024	53	48
March 2024	117	71
April 2024	139	60
May 2024	161	68
June 2024	126	32
July 2024	209	49

Heat maps are visual tools that show data density in geographic areas. In citizen security, areas with the highest frequency of incidents are indicated using a color scale. This helps the authorities to identify “hot spots” for better allocation of resources and preventive measures. Fig. 11 and Fig. 12 heat maps identified critical areas in Cayma (La Tomilla Zone B and Señorial Urbanization) and José Luis Bustamante y Rivero (Estados Unidos Avenue and Vidaurrazaga Avenue), enabling efficient targeting of resources.

Table II presents 413 incidences received, the most common were family violence and alcoholism, with 148 and 144 cases respectively, facilitating a specific preventive approach. These results demonstrate Smart Muni's effectiveness in improving community safety and well-being through effective communication and comprehensive resource management.

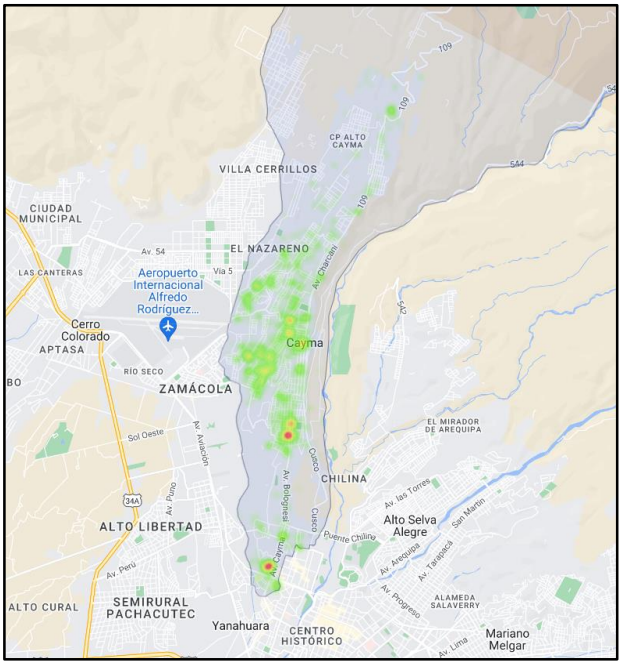


Fig. 11. Heat map of Cayma district.

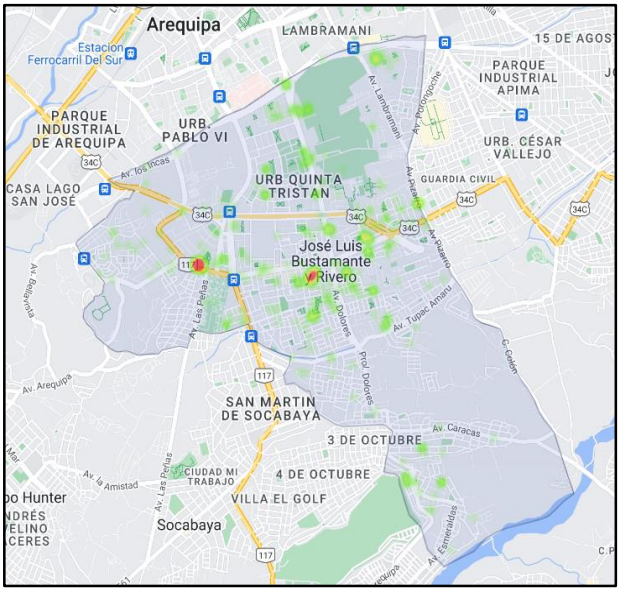


Fig. 12. Heat map of José Luis Bustamante y Rivero district.

TABLE II. NUMBER OF ALERTS AND INCIDENTS BY TYPE

Type of incident	Quantity	Percentage
Family violence	148	36%
Alcoholism	144	35%
Robbery	28	7%
Drug addiction	7	2%
Vandalism	7	2%
Accident	8	2%
Other	71	17%

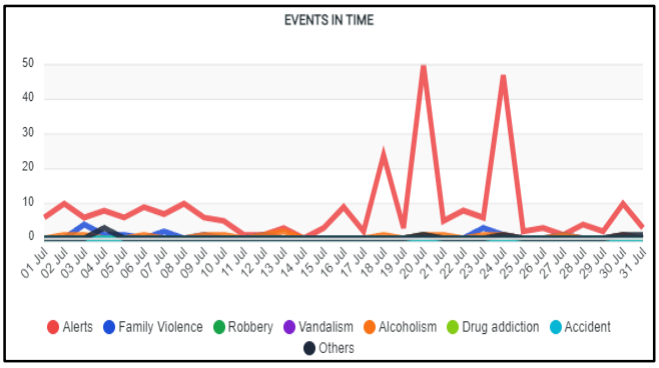


Fig. 13. Alerts and incidents graph in month with the highest activity.

Fig. 13 shows a graph illustrating the quantity of alerts and incidents recorded during the last year. The horizontal axis shows the dates, while the vertical axis shows the quantity of alerts received. Data analysis reveals the day with the highest number of recorded alerts reached a peak of 50, while the daily average number of alerts remained at 10. This graph displays trends and peaks in alert activity, providing a clear view of peak demand times and helping to plan a better response by authorities.

Fig. 14 shows a distribution graph of incident types reported through Smart Muni platform. This graph indicates that the most frequent incidence types are alcoholism and family violence, which stand out significantly over other incidence types. This information is crucial for authorities, helping to identify the areas of greatest concern and need for intervention, facilitating implementation of targeted strategies to deal more effectively with these specific problems.

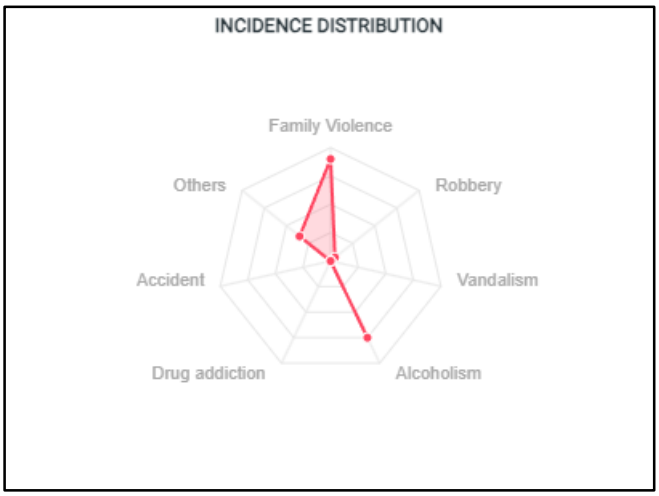


Fig. 14. Distribution of incidence types.

V. Discussion

The implementation of Smart Muni has proven to be an effective solution for emergency management and improving citizen security in the districts of Cayma and José Luis Bustamante y Rivero. The proposed platform, Smart Muni, as suggested by Liu and Li [13], uses geolocation and real-time communication technologies to enhance coordination between citizens and authorities. The implementation of tools such as the digital panic button and instant notifications has significantly reduced emergency response times, in line with recommendations of Liu and Li's study.

Smart Muni incorporates multiple layers of redundancy and uses cloud services such as Firebase and Amazon S3 to ensure data availability and reliability. Although this approach has proven to be effective, it is essential to continue to explore and improve the resilience of the system to potential failures, as suggested by Jesus et al. [14].

While Smart Muni shares similarities with several solutions reviewed by Costa et al. [15], particularly in the use of mobile applications for alert management, it offers significant advantages over existing systems. First, Smart Muni is designed to be highly scalable and flexible, making it adaptable to both large and small municipalities with varying levels of technological infrastructure. Unlike many international systems that are often cost-prohibitive for smaller cities, Smart Muni provides a cost-effective solution that is particularly suited to municipalities operating under budget constraints. Additionally, its integration of real-time communication through services such as Firebase and Twilio ensures a faster, more reliable response compared to traditional systems, which often suffer from delays due to manual processes. Moreover, the cloud-based architecture ensures high availability and data redundancy, critical for maintaining system functionality even in the face of failures, a challenge that many other systems do not adequately address. These advantages make Smart Muni an ideal solution for developing regions where emergency response systems are often underfunded and technologically underdeveloped.

One of Smart Muni's strengths is the implementation of a flexible and scalable architecture that quickly adapts to changing community needs and emerging challenges. This flexibility and adaptability are aligned with the recommendations of Samarakkody et al. [16], who emphasize the importance of dynamic and resilient systems to efficiently manage emergencies in complex urban environments.

Comparison with these previous studies has identified areas for improvement and opportunities for future research. Although Smart Muni has shown promising results in terms of efficiency and responsiveness, it is essential to continue optimizing the system's resilience to failures and exploring new technologies which can complement and enhance existing functionalities. Additionally, performing periodic evaluations and adhering to best practices identified in the literature is recommended to ensure the system remains effective and relevant.

VI. Conclusion

Smart Muni has demonstrated to be a viable and effective solution to improve citizen security through use of advanced technologies. The platform has optimized emergency management and has also strengthened the relationship between citizens and authorities, promoting a safer and more resilient community.

Smart Muni application was successfully developed and implemented, offering key functionalities such as incident reporting, geolocation, real-time notifications and digital panic button; these features have allowed for a more agile and efficient management of emergencies.

Multiple technologies were integrated, including Firebase, Amazon S3, Docker, and Twilio, among others, to guarantee a scalable and highly available system. This technological integration has permitted platform to handle large volumes of data and support fluid communication between different actors involved.

The platform has provided advanced analysis and monitoring tools, which have made it possible to identify critical areas and incidence patterns. This has facilitated a more efficient allocation of resources and implementation.

The work performed provided a solid base for future improvements and expansions of system:

1) Integration of security cameras into platform to enable real-time monitoring. This will allow authorities to view streaming cameras when a robbery is reported, ensuring a faster and more effective response.

2) Implement an option to make payments directly through the application, such as payment of municipal taxes and other municipal services. This will simplify the process for citizens and improve administrative efficiency.

3) Incorporate an intelligent chatbot to help solve users' doubts about use of application and other related issues. This will improve user experience and reduce workload for support staff.

4) Perform data analysis to evaluate security in areas within a 5 km radius, based on reports from the last 2 to 3 weeks. This will allow citizens to have a clear view of security in their area and make informed decisions.

5) Implement a function like a Facebook wall where citizens can interact anonymously, post incidents and respond to municipality publications. This will encourage greater citizen participation and improve communication between community and authorities.

6) Implement a messenger chat so the municipality can have direct communication with neighborhood councils, who are people in charge of supporting citizen security in a specific sector. This will improve coordination and response to emergency situations and security problems.

Despite its success, Smart Muni faced several challenges during its implementation. One of the main limitations was the need for continuous optimization of the system's resilience to potential failures, particularly in areas with unstable internet connectivity, which could affect real-time communication between citizens and authorities. Additionally, the platform's reliance on cloud-based services, while advantageous for scalability, also presents a challenge in terms of maintaining high availability in regions with limited infrastructure. Future improvements should focus on enhancing the system's ability to operate efficiently under these conditions and ensuring its long-term sustainability in municipalities with constrained budgets. Addressing these challenges will be critical for further expanding the platform's reach and impact in other regions.

ACKNOWLEDGMENT

The authors express their gratitude to the Universidad Nacional de San Agustin of Arequipa for the support provided in the realization of this project. The municipalities of Cayma and José Luis Bustamante y Rivero of the city of Arequipa-Peru are also thanked for their collaboration in the implementation of Smart Muni platform. Finally, acknowledgement is extended to all citizens who actively participated in testing and validation of the system.

REFERENCES

[1] J. Núñez, "Fernando Carrión (ed.) (2002). Seguridad ciudadana, ¿espejismo o realidad? Quito: FLACSO Ecuador - OPS/OMS," EURE (Santiago), vol. 29, no. 88, pp. 179-177, 2003. DOI: https://dx.doi.org/10.4067/S0250-71612003008800008.

[2] M. García Ojeda and A. Zambrano Constanzo, "Seguridad ciudadana: el aporte de las metodologías implicativas," Revista de Psicología, vol. 14, no. 2, pp. 63-79, 2005.

[3] A. Olmo, "Desarrollo e implementación de un simulador de datos GNSS," Ph.D. dissertation, Univ. de Jaén, Jaén, 2016.

[4] Instituto Nacional de Estadística e Informática, "Encuesta Nacional de Programas Estratégicos," INEI, Lima, Peru, 2017.

[5] A. Sarkar, "Technology and Innovations: Empowering Citizens for Future Liveable," International Journal of Management and Humanities, vol. 4, no. 11, pp. 16-22, 2020.

[6] A. Samarakkody, D. Amaratunga, and R. Haigh, "Technological Innovations for Enhancing Disaster Resilience in Smart Cities: A Comprehensive Urban Scholar's Analysis," Sustainability, vol. 15, no. 15, Article ID 12036, 2023. DOI: https://doi.org/10.3390/su151512036.

[7] K. Borsekova, P. Nijkamp, and P. Guevara, "Urban resilience patterns after an external shock: An exploratory study," International Journal of Disaster Risk Reduction, vol. 31, pp. 381–392, 2018.

[8] W. Castelnovo, G. Misuraca, and A. Savoldelli, "Smart Cities Governance: The Need for a Holistic Approach to Assessing Urban Participatory Policy Making," Social Science Computer Review, vol. 34, no. 6, pp. 724-739, 2015. DOI: https://doi.org/10.1177/0894439315611103.

[9] D. Icaza Álvarez, "EJE 07-10 Sistemas de seguridad ciudadana por georeferenciación y geolocalización para zonas rurales del cantón Cuenca incorporados al SIS ECU 9-1-1 del Ecuador," Memorias Y Boletines De La Universidad Del Azuay, vol. 1, no. XVI, pp. 413–418, 2017. DOI: https://doi.org/10.33324/memorias.v1iXVI.88.

[10] A. Espinoza-Ramírez, M. Nakano, G. Sánchez-Pérez, and A. Arista-Jalife, "Sistemas de Información Geográfica y su Análisis Aplicado en Zonas de Delincuencia en la Ciudad de México," Información tecnológica, vol. 29, no. 5, pp. 235-244, 2018. DOI: https://dx.doi.org/10.4067/S0718-07642018000500235.

[11] M. González, "Sobre la inconstitucionalidad de la geolocalización de los ciudadanos mediante el uso de aplicaciones de dispositivos móviles al amparo de la Orden SND 297/2020," Diario La Ley, no. 9643, 2020.

[12] A. Solano-Barliza, "Revisión conceptual de sistemas de recomendación y geolocalización aplicados a la seguridad turística," Journal of Computer and Electronic Sciences: Theory and Applications, vol. 2, no. 2, 2021. DOI: 10.17981/cesta.02.02.2021.05.

[13] H. Liu and Y. Li, "Smart cities for emergency management," Nature, vol. 578, no. 7796, pp. 515-516, 2020.

[14] T. C. Jesus, P. Portugal, D. G. Costa, and F. Vasques, "Reliability and detectability of emergency management systems in smart cities under common cause failures," Sensors, vol. 24, no. 9, p. 2955, 2024.

[15] D. G. Costa, J. P. J. Peixoto, T. C. Jesus, P. Portugal, F. Vasques, E. Rangel, and M. Peixoto, "A survey of emergencies management systems in smart cities," IEEE Access, vol. 10, pp. 61843-61872, 2022.

[16] A. Samarakkody, D. Amaratunga, and R. Haigh, "Technological innovations for enhancing disaster resilience in smart cities: a comprehensive urban scholar's analysis," Sustainability, vol. 15, no. 15, p. 12036, 2023.

[17] Municipalidad Distrital de José Luis Bustamante y Rivero. Available: https://www.munibustamante.gob.pe/archivos/1647009631.pdf. Accessed May 23, 2024.

[18] Municipalidad Distrital de Cayma. Available: https://www.municayma.gob.pe/wpfd_file/formulacion-del-plan-de-accion-distrital-de-seguridad-ciudadana-cayma-2023/. Accessed May 23, 2024.

[19] Amazon S3, Amazon Web Services. Available: https://aws.amazon.com/s3/. Accessed May 23, 2024.

[20] PostgreSQL, PostgreSQL Global Development Group. Available: https://www.postgresql.org/. Accessed May 23, 2024.

[21] Firebase Realtime Database, Google Firebase. Available: https://firebase.google.com/products/realtime-database. Accessed May 23, 2024.

[22] Docker, Docker Inc. Available: https://www.docker.com/. Accessed May 23, 2024.

[23] Docker Compose, Docker Inc. Available: https://docs.docker.com/compose/. Accessed May 23, 2024.

[24] GitLab CI/CD, GitLab Inc. Available: https://about.gitlab.com/features/gitlab-ci-cd/. Accessed May 23, 2024.

[25] GitLab Runner, GitLab Inc. Available: https://docs.gitlab.com/runner/. Accessed May 23, 2024.

[26] DigitalOcean, DigitalOcean LLC. Available: https://www.digitalocean.com/. Accessed May 23, 2024.

[27] GoDaddy, GoDaddy Inc. Available: https://www.godaddy.com/. Accessed May 23, 2024.

[28] Firebase Cloud Messaging, Google Firebase. Available: https://firebase.google.com/products/cloud-messaging. Accessed May 23, 2024.

[29] Twilio, Twilio Inc. Available: https://www.twilio.com/. Accessed May 23, 2024.

[30] Django REST Framework, Encode OSS Ltd. Available: https://www.django-rest-framework.org/. Accessed May 23, 2024.

[31] Quasar Framework, Quasar Framework. Available: https://quasar.dev/. Accessed May 23, 2024.

[32] Flutter, Google LLC. Available: https://flutter.dev/. Accessed May 23, 2024.

[33] Nginx, NGINX Inc. Available: https://www.nginx.com/. Accessed May 23, 2024.

[34] uWSGI, uWSGI Project. Available: https://uwsgi-docs.readthedocs.io/. Accessed May 23, 2024.

[35] Let's Encrypt, Internet Security Research Group. Available: https://letsencrypt.org/. Accessed May 23, 2024.

[36] Ubuntu UFW, Canonical Ltd. Available: https://help.ubuntu.com/community/UFW. Accessed May 23, 2024.

Serious Games Model for Higher-Order Thinking Skills in Science Education

Siti Norliza Awang Noh[1], Hazura Mohamed[2], Nor Azan Mat Zin[3]
Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi Selangor, Malaysia

***Abstract*—The popularity of digital games has led to the emergence of serious games, which are developed with specific purposes beyond mere entertainment. Serious games in education represent more innovative and current pedagogical approaches. However, the existing digital games have been shown to improve critical thinking skills, although there is still a limited amount of research on science education. A preliminary study has found that digital games developed for science teaching do not incorporate all aspects of Higher-Order Thinking Skills (HOTS). This study aims to identify and validate game components and design a serious game model for HOTS in science education (PKBATDPS Model), which was validated using the Electric Circuit prototype. The study is divided into four phases: analysis, design, development and evaluation. During the analysis phase, the components of the PKBATDPS model were identified. The Electric Circuit prototype was evaluated using a quasi-experimental procedure that included pre-tests, post-tests, and learning motivation questionnaires. The experiment involved 32 elementary students; 16 in the experimental group used the serious games application prototype, whereas 16 in the control group received the traditional method. The results show that the PKBATDPS Model can be effectively used to increase students' HOTS and motivation in science education.**

***Keywords*—Serious game; Higher-Order Thinking (HOT) skills; science education; game element; learning element**

I. INTRODUCTION

Serious games are interactive digital games designed with specific purposes other than entertainment that can train and educate players [1][2]. In recent years, serious games have been developed and used in multiple fields, including the military, advertising, education, simulation, and healthcare [3][4]. In the field of education, serious games are one of the technological tools used as a teaching aid to meet the needs of the current generation. This generation is born in the age of digital technology, commonly known as the Digital Native generation, and is already familiar with technology [5]. Today's generation prefers active learning that is interactive, enjoyable, and focused on problem solving. Therefore, the use of serious games as teaching aids is in line with educational objectives, as they are designed to provide an interactive and exciting learning experience.

According to [6], serious games are a form of student-centred approach that have the potential to build students' understanding of learning. Serious games have grown significantly and widely used in many different subjects to support the learning process and provide learning experiences such as in mathematics [7], science [8], history [9] and language [10]. The use of serious games in education has been shown to have the potential to foster an engaging and interactive learning environment that cultivates critical thinking skills among students. This is because serious games involve the development of the mind and require a deeper level of thinking [6].

Hence, the integration of serious games in education should be broadened to all academic disciplines, with particular emphasis on science. It is regarded as challenging due to its abstract nature and Higher Order Thinking Skills (HOTS) presence. HOTS refers to the highest level of cognitive processing in the cognitive function hierarchy. HOTS entails receiving, storing, delivering, and synthesising new information with existing knowledge to resolve intricate problems [11]. Regarded as pivotal competencies, HOTS contribute to shaping a generation equipped with innovative thinking and the requisite abilities to confront the demands of the 21st century [12].

In Malaysia, the incorporation of thinking skills into the curriculum began in the early years of 1993, focusing on promoting critical and creative thinking skills [13][14]. In 2011, the curriculum was further enhanced to include HOTS, to stimulate students' thinking skills [13]. However, issues related to ineffective science teaching methods have hindered the successful implementation of the HOTS [15]. Traditional teaching methods, which emphasise cognitive activities such as memorisation, recall, and comprehension, have been deemed inadequate to effectively implement HOTS [12]. One-way communication learning methods, such as "note-taking and lecturing," are also viewed as teacher-centred instructional approaches. This method restricts students' ability to think critically and generate and develop their ideas, particularly in science education.

Currently, research on the design of games for science education is also being conducted. According to [16][17], serious games emerge as an effective tool that offer engaging and enjoyable experiences for enhancing learning in science. This is because serious games offer students the opportunity to engage in other alternatives to learning in simulation or experiments [18][19]. Moreover, serious games provide a more active and interactive learning experience that can stimulate interest, increase motivation [2], increase engagement [16] and improve academic achievement in science [20].

However, one of the identified issues in serious game design is the weakness in terms of learning content. In the design of games for science education, alignment with local

curriculum content is imperative to meet the requirements of the national education system and the students' needs. This necessity arises due to disparities in curriculum content, which render existing games on the market less suitable for instructional purposes within the Malaysian context [21].

Previous research has investigated digital games to enhance critical thinking skills [22], yet there remains a scarcity of studies specifically focusing on science education [23]. Preliminary investigations have revealed that digital games designed for science education often lack the integration of all components of HOTS. Additionally, the content of these digital games frequently fails to establish connections with real-life scenarios [21]. Therefore, the collaboration between game developers and subject matter experts is needed to ensure that the serious game design is in line with the learning content that emphasises the HOTS in science education. Hence, developing a model for HOTS games in science education is needed to captivate students' interest and enhance their HOTS capabilities within the subject matter.

II. METHOD

This study utilized the Design and Develop Research (DDR) approach, which encompasses four primary phases: analysis, design, development and evaluation. Data collection employed a mixed-methods approach, integrating both qualitative and quantitative research methodologies.

A. Analysis Phase

The analysis phase serves as the initial step in the study to ascertain the research requirements. During this phase, issues and problems, along with elements of the game model, were identified. The activities include a comprehensive literature review, a survey, and unstructured interviews. The literature explored students' perceptions of science education, the challenges of implementing HOTS in science education, and the design of serious games. The survey aimed to capture students' perceptions of the science subject. Unstructured interviews were conducted with a panel of eight experts, comprising six science teachers and two School Improvement Specialist Coaches Plus (SIC+) science officers, to identify HOTS implementation in science education and the role of digital games in educational settings. Fig. 1 summarizes the activities performed in this phase.

B. Design Phase

In this phase, four key activities were involved: component grouping, component mapping, low-fidelity prototype design, and initial component validation. All elements were systematically organized and mapped into different components of the PKBATDPS Model. The initial PKBATDPS Model was categorized into two primary components: learning components and game components. To illustrate the interrelationship between these components, the research model was depicted in diagram form, providing a comprehensive overview.

Subsequently, the components of the PKBATDPS Model in the low-fidelity prototype, namely the Electric Circuit storyboard, were implemented. The activities are developed based on the Standard Curriculum and Assessment Document for Year 5 Science, specifically the electric topic outlined in the Primary School Standard Curriculum. This approach ensures that the prototype is grounded in the existing educational framework and aligns with the instructional resources available to students.

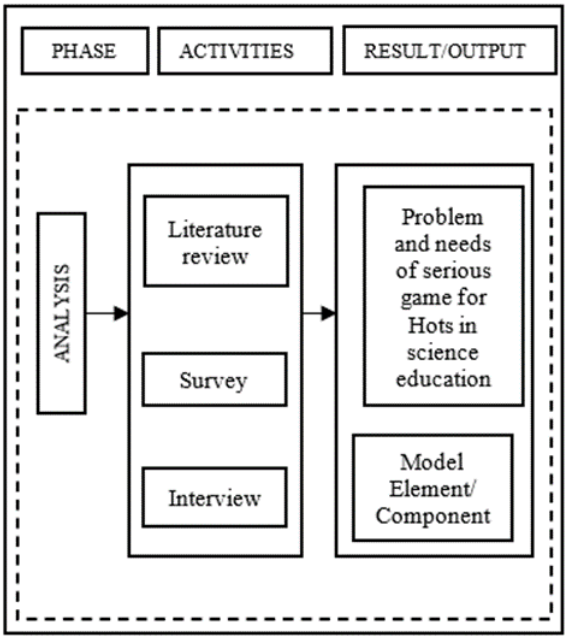


Fig. 1. The activities in the analysis phase.

The initial validation of the PKBATDPS Model components was conducted through heuristic evaluation by assessing the low-fidelity prototype. The validation of the PKBATDPS Model elements engaged two SISC+ officers from the District Education Office (DEO) and three university lecturers. The number of expert evaluators selected for this evaluation is aligned with the recommendations of [24], which advocate for a panel size of three to five individuals. The evaluators were chosen based on their expertise and experience in science education, as well as their proficiency in software and game assessment. Expert validation was conducted to ascertain the implementation of each model element in the game interface design.

The questionnaire consists of three sections: Demographic Information, Learning, and Game. Each expert evaluator responds to questions using a five-point Likert scale, where 1 indicates "Strongly Disagree", 2 “Disagree”, 3 “Neutral”, 4 “Agree” and 5 indicates "Strongly Agree". The assessment results are analysed using the Interquartile Range (IQR) to determine consensus in evaluating expert opinions [25][26]. The IQR is a measure of variability based on the distribution of data divided into three parts: the first quartile (Q1), the median or second quartile (Q2) and the third quartile (Q3). The IQR value is calculated by subtracting Q1 from Q3. Elements with an IQR value of 1 or less (≤ 1) are retained, while those with IQR value greater than (>1) are eliminated. Fig. 2 provides a summary of the activities performed in this phase.

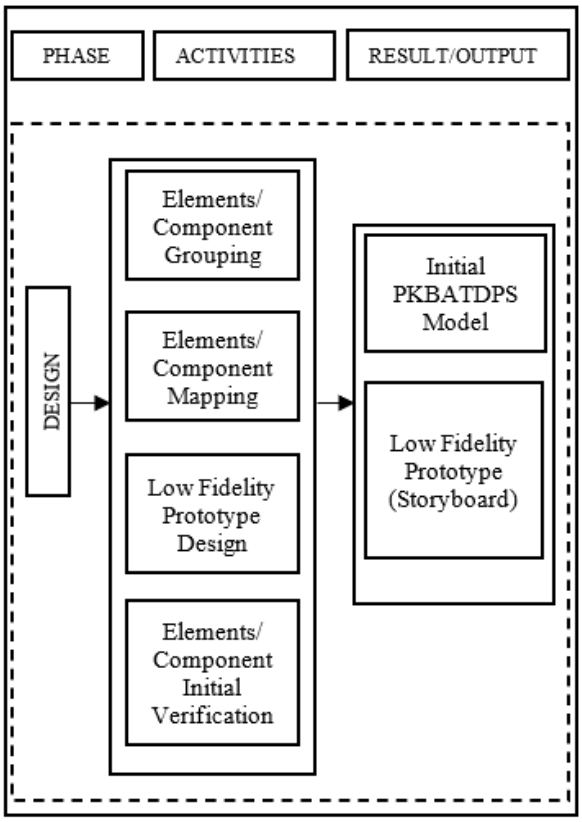


Fig. 2. The activities in the design phase.

C. Development Phase

The development phase involves the implementation of the components of the PKBATDPS Model, which have been refined based on feedback from experts, into the high-fidelity prototype, namely the Electric Circuit game prototype. The development of the Electric Circuit game prototype utilized three software applications: Adobe Photoshop, Autodesk 3D Max, and UNITY, which were employed in the process of creating the game interface. Subsequently, the development of modules, quizzes, and games was executed based on the content design established during the design phase.

The Electric Circuit game prototype is software played offline and is based on self-directed learning, allowing students to learn according to their convenience in terms of place and time.

The game prototype encompasses various subtopics, including electric power sources, complete electric circuits, and safety measures for handling electrical equipment, all presented within a narrative framework. The developmental process of the prototype unfolds progressively through different levels of the game. Advancement from one level to the next entails activities designed to cultivate HOTS, such as application, analysis, evaluation, and creation. This pedagogical approach mirrors Bloom's taxonomy (1985), which underscores the necessity for learners to master each instructional unit before progressing to subsequent ones.

Finally, a meeting was convened with the science subject teachers to provide a concise overview of the study's objectives, evaluation methods, and the prototype employed. During this session, teachers received training on the operational aspects of the Electric Circuit game prototype prior to the actual implementation and evaluation. This preparatory measure ensured that teachers were well-equipped to facilitate the learning experience and offer constructive feedback during the assessment phase, thereby enhancing the overall efficacy of the evaluation process. Fig. 3 shows the activities conducted during this phase.

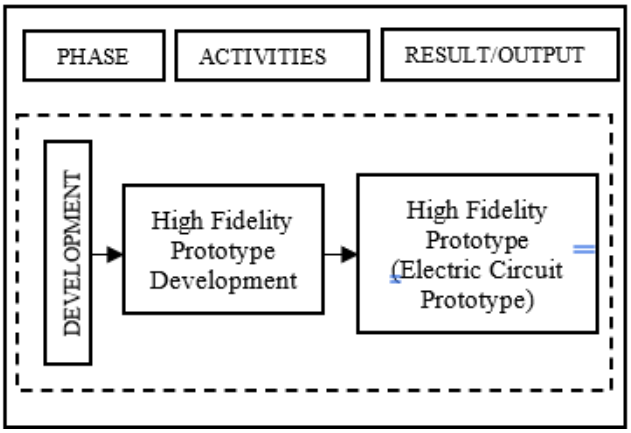


Fig. 3. The activities in the development phase.

D. Evaluation Phase

The implementation and evaluation phases involve three activities: Instrument Development, Content validation and Pilot Study, and Evaluation of the Electric Circuit game prototype. The evaluation process involves 32 fifth-grade students, divided into control and experimental groups. Participants are selected based on equivalent achievement scores from previous science assessments. This aligns with the recommendations suggesting a suitable sample size of 10 to 20 participants for comparative experimental studies [27][25]. Additionally, a minimum sample size of 15 participants is recommended for quantitative experimental research [28][29].

The independent variable in this study is the teaching method, specifically the use of serious games compared to traditional instructional methods. The dependent variables are students' scores on the science HOTS test and their motivation survey scores in science. The evaluation instrument is divided into two types: pre- and post-tests (and a motivation survey). The pre-test comprises a series of questions on the electric topic in science, generated based on the Primary School Standard Curriculum syllabus, textbooks, and reference materials for fifth-grade science. This test aims to assess students' HOTS levels in science before and after exposure to serious games.

The test instruments consist of two sections: Demographic Information and science questions which are categorised into four types. The test instruments were developed with the assistance of several teachers and subsequently reviewed by two subject-matter experts skilled in the relevant field. This expert review is essential to ensure the accuracy of the constructs and the clarity of the research instrument content, as recommended by [30].

The motivation questionnaire was then developed based on Keller's (1987) ARCS model, which includes the elements of Attention, Relevance, Confidence, and Satisfaction. This instrument was created to assess students' motivation levels before and after science learning using serious games. The questionnaire comprises two sections: Section A gathers demographic information, while Section B includes four constructs and 27 items measured using a 5-point Likert scale to evaluate student motivation in science.

A pilot study involving 15 randomly selected fifth-grade students was conducted to evaluate the reliability of a motivation questionnaire within the context of science education. The data collected were analyzed using Cronbach's alpha, following Cronbach's (1946) methodology, with IBM SPSS Statistics version 16.0. The overall reliability score of the instrument was 0.81, indicating a high level of reliability. This result suggests that the motivation questionnaire is suitable for use in actual research [31][32]. Fig. 4 illustrates the activities during the implementation and evaluation phases.

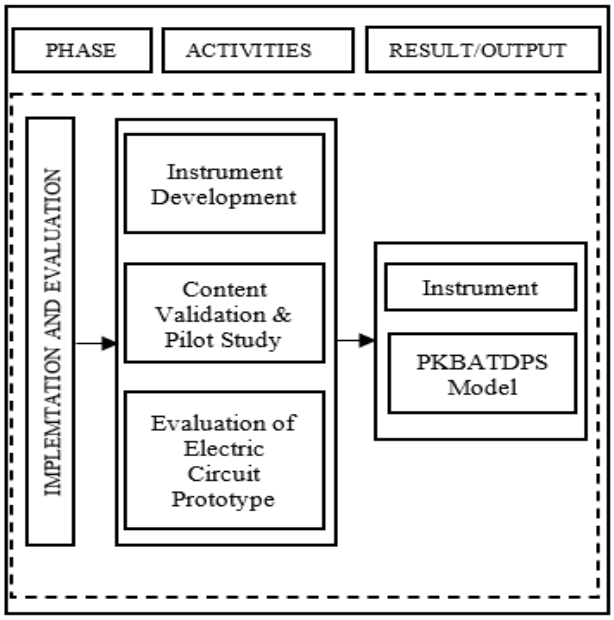


Fig. 4. The activities in the evaluation phase.

III. Results and Discussion

This study successfully identified 10 game elements categorised into two main components in the PKBATDPS Model. These ten elements consist of four learning elements and six game elements. The learning elements consist of learning objectives, constructivism learning theory, science subject content emphasising HOTS, and the ARCS motivation model. Meanwhile, the game elements consist of objectives, fantasy, challenge, feedback, control, and rule. One for both learning and game components. This indicates that the experts agreed that all assessed components of the PKBATDPS Model, as depicted in Fig. 5, were effectively implemented in this low-fidelity prototype.

Therefore, no elements needed to be eliminated or added. However, the experts provided feedback and suggestions for improving the design of the Electric Circuit prototype, particularly concerning the implementation of goal elements. They recommended that the game goal element be more clearly indicated at the end of the game to show that the game has successfully achieved the stated goal presented at the beginning. Improvements were incorporated during the development of the Electric Circuit prototype. Table I presents the learning and game elements validated by the experts, along with their IQR values.

TABLE I. PKBATDPS Model Elements Validated by Experts (IQR Value)

Bil	Learning Element	P1	P2	P3	P4	P5	Q1	Q3	IQR
1.	Learning Goal	5	5	4	5	4	4	5	1
2.	Constructivism Learning Theory	4	4	5	5	4	4	5	1
3.	Science Learning Content (Based on HOTS Element)	4	4	5	4	5	4	5	1
4.	ARCS Model Motivation	4	5	4	5	4	4	5	1
	Game Element	**P1**	**P2**	**P3**	**P4**	**P5**	**Q1**	**Q3**	**IQR**
1.	Goal	5	5	4	4	5	4	5	1
2.	Fantasy	5	4	5	4	4	4	5	1
3.	Challenge	5	5	4	4	4	4	5	1
4.	Feedback	5	5	4	4	5	4	5	1
5.	Control	5	5	4	4	5	4	5	1
6.	Rule	5	4	5	5	4	4	5	1

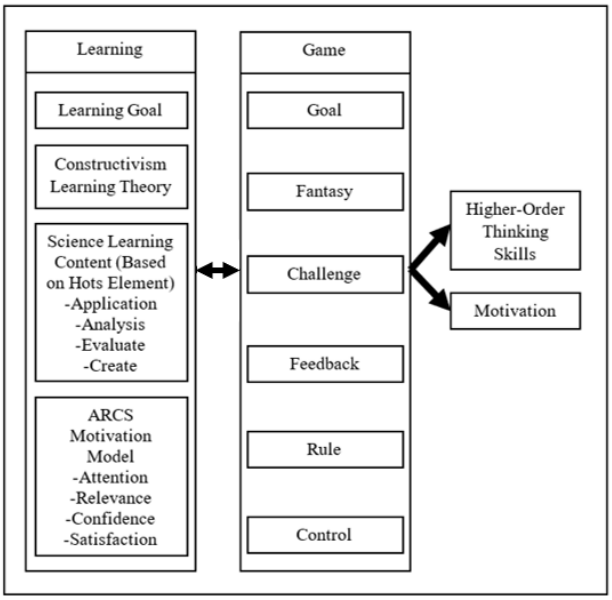


Fig. 5. PKBATDPS model.

The evaluation of the Electric Circuit prototype was conducted using a quasi-experimental method, which included pre-tests, post-tests, and learning motivation questionnaires. Two groups were involved: a control group and an experimental group.

E. Analysis of HOTS

The result revealed a significant difference in the mean scores of pre-tests and post-tests for the experimental group. Table II shows the mean score for pre-test of 16.38 and post-test of 64.63. These results indicate that there is an improvement in HOTS among students after using serious game applications in science education.

TABLE II. T-TEST RESULT OF THE PRE AND POST-TESTS FOR EXPERIMENTAL GROUP

Experiment Group	N	Mean	Standard Deviations	t	p
Pre-Test	16	16.38	9.13	16.15	0.000
Post-Test	16	64.63	9.95		

Next, the result indicated a significant difference in the mean post-test scores between the experimental and control groups. Table III shows the mean post-test score for the experimental group was 64.63, higher than the control group's mean post-test score of 29.38. These results indicated that the experiment group who engaged with the serious game application demonstrated better HOTS than the control group who received the traditional method in science.

TABLE III. T-TEST RESULT OF THE POST-TEST SCORE FOR THE EXPERIMENTAL AND THE CONTROL GROUP

Sample	Mean	Standard Deviations	t	p
Experiment Group	64.63	9.95	11.01	.000
Control Group	29.38	8.05		

F. Analysis of Motivation

The result demonstrated a significant difference in the mean scores of motivation before and after using serious game applications in science education. Table IV shows that the mean motivation score after using serious game applications was 4.45, higher than before using serious game applications. These results indicated that the students' motivation increased after using serious game applications in science education.

TABLE IV. TEST RESULT OF THE LEARNING MOTIVATION POST-QUESTIONNAIRE RATING FOR EXPERIMENTAL GROUP

Experiment Group	N	Mean	Standard Deviations	t	p
Motivation-Before	16	4.00	.30	5.39	0.000
Motivation-After	16	4.45	.28		

Then, the motivation result between the two groups indicated a significant difference in the mean motivation score. Table V shows that the mean motivation score for the experiment group was 4.45 higher compared to the control group mean score of 4.00. These results indicate that students who used the serious game application prototype in science education demonstrated higher motivation compared to those who learnt through traditional methods.

TABLE V. T-TEST RESULT OF THE LEARNING MOTIVATION POST-QUESTIONNAIRE RATING FOR EXPERIMENTAL GROUP AND CONTROL GROUP

Sample	Mean	Standard Deviations	t	p
Experiment Group	4.45	.28	3.98	.000
Control Group	3.95	.41		

Overall, these assessment results demonstrate that using the Electric Circuit prototype effectively enhances Higher-Order Thinking Skills (HOTS) and student motivation in science education. Consequently, it can be concluded that the PKBATDPS Model, as implemented in the Electric Circuit prototype, can be utilized to cultivate students' HOTS in science education. Additionally, it has the potential to increase students' motivation and interest in science education.

IV. CONCLUSION

This study successfully identified 10 elements categorised into two main components in the PKBATDPS Model and implemented in Electric Circuit prototype. The evaluation of the Electric Circuit prototype was conducted using a quasi-experimental method, which included pre-tests, post-tests, and learning motivation questionnaires. Two groups were involved: a control group and an experimental group. The experimental group participated in learning activities using the serious game application while the control group engaged in traditional learning method. The results show that students in the experiment group demonstrated a significant improvement in HOTS after using serious game applications in science education. In addition, the results reveal that the student in the experiment group who engaged with the serious game application demonstrated better HOTS than the student in the control group who received the traditional method in science. Observations showed that the students in the experimental group enjoy playing games with science. In terms of learning motivation, the result shows that the students' motivation in experiment group increased after using serious game application in science education. Furthermore, the result indicate that students who used the serious game application prototype in science education demonstrated higher motivation compared to those who learned through traditional methods. These results indicate that the implementation of serious game application in learning provides more engaging and interactive learning environment that cultivate HOTS and increase motivation in science education. Therefore, serious games can be considered a novel learning method that is more engaging, interactive, and effective in enhancing HOTS and student motivation in science education. The evaluation results of the Electric Circuit game prototype validate the proposed PKBATDPS Model suggested by this study. The results demonstrate that the objectives of the study have been achieved. However, further research can be carried out by adding learning elements and other game elements to challenge the player's level of thinking at a high level.

ACKNOWLEDGMENT

Acknowledge Sekolah Kebangsaan Othman Talib Dua, Sepang District Education Office and Pasir Mas District Education Office. The Malaysian Ministry of Education for funding the under program MyBrain15 (MyPhD). The authors also would like to acknowledge and extend special gratitude to the Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, who granted the Publication Grant Scheme for this project.

REFERENCES

[1] C. A. D.León, N. A. Rivera, M. G. Cabrera, J. H. M. Cano, S. A. Ortiz, and J. D. M. Osorio, "Designing learning experiences using serious games: innovative village case study," Front. Educ., vol. 9, no, 2024.

[2] Y. M. Arif, N. Ayunda, N. M. Diah, and M. B. Garcia, "A systematic review of serious games for health education: Technology, challenges, and future directions," Transform. Approaches to Patient Lit. Healthc. Innov., pp. 20–45, 2024.

[3] M. Mansor, R.A. Wahid, S. Sulaiman, A. H. Ariffin, and N. Azme, "WhysoSeriousGame: A conceptual framework for learning motivation through serious game," J. ICT Educ., vol. 8, no. 4, pp. 122–129, 2021.

[4] H.. Xin and N. Mohamad Ali, "Serious Games Technology for Older Adults," in Interaksi Proceeding 2023.

[5] C. Wang and L. Huang, "A Systematic Review of Serious Games for Collaborative Learning: Theoretical Framework, Game Mechanic and Efficiency Assessment," Int. J. Emerg. Technol. Learn., vol. 16, no. 6, pp. 88–105, 2021.

[6] M. Ahmad, N. R. Mansor, R. A. Rashid, N. A. Chua, R. Zakaria, and C. M. Sung, "Implementation of Digital Games in Advancing Students' Higher-Order Thinking Skills: A Review," J. Phys. Conf. Ser., vol. 1793, no. 1, 2021.

[7] I. Sarifah, A. Muhajir, A. Marini, G. Yarmi, D. Safitri, and L. Dewiyani, "Mobile games and learning interest: for fifth graders in mathematics," J. Educ. Learn., vol. 19, no. 1, pp. 151–157, 2024.

[8] M. Ullah et al., "Serious games in science education : a systematic literature review," Virtual Real. Intell. Hardw., vol. 4, no. 3, pp. 189–209, 2022.

[9] C. Y. Chong, "The Effectiveness of Digital Game-Based Learning (DGBL) on Primary School Students' Achievement in History The Effectiveness of Digital Game-Based Learning (DGBL) on Primary School Students' Achievement in History," Innov. Teach. Learn. J., vol. 5, no. 1, pp. 49–53, 2021.

[10] K. Ishaq, F. Rosdi, N. A. Mat Zin, and A. Abid, Serious game design model for language learning in the cultural context, vol. 27, no. 7. Springer US, 2022.

[11] L. Luzyawati, I. Hamidah, A. Fauzan, and H. Husamah, "Higher-order thinking skills-based science literacy questions for high school students," J. Educ. Learn., vol. 19, no. 1, pp. 134–142, 2024.

[12] L. C. Tong, M.S. Rosli, and N. S. Saleh, "Enhancing HOTS using Problem-Based Learning and Digital Game in the Context of Malaysian Primary School," Int. J. Interact. Mob. Technol., vol. 16, no. 2, pp. 101–112, 2022.

[13] C. W. Ting, S. Surat, and S. Rahman, "Tahap Efikasi Kendiri dan Penguasaan Kemahiran Berfikir Aras Tinggi (KBAT) dalam kalangan Murid," J. Dunia Pendidik., vol. 4, no. 4, pp. 318–333, 2023.

[14] S. N. Awang Noh, H. Mohamed, and N. A. Mat Zin, "The Effects of Serious Games on Students ' Higher- Order Thinking Skills in Science Education," in 2021 International Conference on Electrical Engineering and Informatics (ICEEI), 2021.

[15] A. S .A. Halim, K Osman, M. S. A. Mohd Aziz, M. F. Ibrahim, and A. A. K Ahmad, "The Competency of Science Teachers in Integrating Higher Order Thinking Skills in Teaching and Learning," J. Phys. Conf. Ser., vol. 1793, no. 1, 2021.

[16] M. Alahmari, M. T. Jdaitawi, A. Rasheed, R. E. Hussein, R. Abduljawad, M. Alzahrani and N. Awad, "Trends and gaps in empirical research on gamification in science education: A systematic review of the literature," Contemp. Educ. Technol., vol. 15, no. 3, 2023.

[17] G. M. Chans and M. Portuguez Castro, "Gamification as a strategy to increase motivation and engagement in higher education chemistry students," Computers, vol. 10, no. 10, 2021.

[18] N. Kara, "A systematic review of the use of serious games in science education," Contemp. Educ. Technol., vol. 13, no. 2, pp. 1–13, 2021.

[19] M. Riopel et al., "Impact of serious games on science learning achievement compared with more conventional instruction: an overview and a meta-analysis," Stud. Sci. Educ., vol. 55, no. 2, pp. 169–214, 2020

[20] K. Osman and L. A. Nam, "MyKimDG module: an interactive platform towards development of twenty-first century skills and improvement of students' knowledge in chemistry," Interact. Learn. Environ., 2020.

[21] K. Osman and N. A. Bakar, "Educational computer games for Malaysian classrooms: Issues and challenges," Asian Soc. Sci., vol. 8, no. 11, pp. 75–84, 2012.

[22] C. V. Angelelli, G. M. de C. Ribeiro, M. R. Severino, E. Johnstone, G. Borzenkova, and D. C. O. da Silva, "Developing critical thinking skills through gamification," Think. Ski. Creat., vol. 49, no. August 2022, p. 101354, 2023.

[23] M.-T. Cheng, J.-H.. Chen, S.-J. Chu, and S. Y. Chen, "The use of serious games in science education: a review of selected empirical research from 2002 to 2013," J. Comput. Educ., vol. 2, no. 3, pp. 353–375, 2015.

[24] J. Nielsen, "How to Conduct a Heuristic Evaluation," 1994.

[25] N. E. M. Razali, R. Z. Ramli, H. Mohamed, N. A. M. Zin, F. Rosdi, and N. M. Diah, "Identifying and validating game design elements in serious game guideline for climate change," Heliyon, vol. 8, no. 1, 2022.

[26] U. Sekaran, Research Methods for Business: A Skill Building Approach, 4th ed., vol. 49, no. 4. John Wiley & Sons, Icl., 2003.

[27] H. A. V. D. Gracht, "Consensus measurement in Delphi studies Review and implications for future quality assurance," Technol. Forecast. Soc. Change, vol. 79, no. 8, pp. 1525–1536, 2012.

[28] F. Salam, R. Mailok, and N. Ubaidullah, "The Change of Achievements in Information and Communication Technology Subjects through Project-Based Learning with Scaffolding," Malaysian J. Educ, vol. 40, no. 1, pp. 29–41, 2015.

[29] J. Bacatang, "Implementation of a Quasi-Experimental Study to Identify Early Literacy Skills of Preschool Children," in International Language Conference (ILC) 2014, 2014.

[30] J. H. McMillan and S. Schumacher, Research in Exucation: Evidence Based Inquiry, 7th ed. New Jersey: Pearson,Education Incorporated, 2010.

[31] T. Kline, Psychological Testing : A Practical Approach to Design and Evaluation. California: Sage Publication, 2005.

[32] R. Tap, N. A. Mat Zin, H. M. Sarim, and N. M. Diah, "Creativity Training Model for Game Design," vol. 12, no. 5, pp. 59–66, 2021.

Lung CT Image Classification Algorithm Based on Improved Inception Network

Qianlan Liu
School of Information Engineering, Hunan University of Science and Engineering, Yongzhou, 425199, China

***Abstract*—With the continuous development of digital technology, traditional lung computed tomography medical image processing has problems such as complex images, small sample data, and similar symptoms between diseases. How to efficiently process lung computed tomography image classification has become a technical challenge. Based on this, the Inception algorithm is fused with the improved U-Net fully convolutional network to construct a lung computed tomography image classification algorithm model based on the improved Inception network. Subsequently, the Inception algorithm is compared with other algorithms for performance analysis. The results show that the proposed algorithm has the highest accuracy of 92.7% and the lowest error rate of 0.013%, which is superior to the comparison algorithm. In terms of recall comparison, the algorithm is approximately 0.121 and 0.213 higher than ResNet and GoogLeNct algorithms, respectively. In comparison with other models, the proposed model has a classification accuracy of 98.1% for viral pneumonia, with faster convergence speed and fewer required parameters. From this result, the proposed Inception network based lung computed tomography image classification algorithm model can efficiently process data information, provide technical support for lung computed tomography image classification, and thereby improve the accuracy of lung disease diagnosis.**

Keywords*—Image classification; inception; lung CT images; CNN; machine learning

I. Introduction

In recent years, the rapid development of medical technology has promoted the extensive application of computer-aided diagnosis system in the medical field, especially in the diagnosis of pulmonary diseases, where computer-aided system has shown great potential [1]. As a key link in the diagnosis of lung diseases, the accuracy of lung CT image classification directly affects the early detection and identification of diseases [2]. However, due to the complex structure and rich information of lung CT images, traditional classification methods are often difficult to effectively process, which not only limits the timely diagnosis of lung diseases, but also adversely affects the treatment effect and survival rate of patients [3]. In current studies, although deep learning has made many breakthrough achievements in the field of image classification, more refined and efficient algorithm models are still needed for specific challenges in lung CT image classification, such as high image complexity, limited sample data, and similar inter-disease symptoms [4-5]. Inception, as an important part of deep learning, can capture features of different levels through multiple branch networks and convolution nuclei of different scales, which provides the possibility to process the complex structure and rich information of lung CT images. Although the existing Inception network performs well in image classification, it still has some limitations when processing lung CT images. For example, the low degree of association between feature pixels and target pixels, as well as weak gradients of Inception module information, affect the accuracy and efficiency of classification. Therefore, this study aims to propose a lung CT image classification algorithm based on improved Inception network to enhance feature extraction and classification performance by fusing full convolutional networks (such as U-Net) to solve the problems encountered by existing algorithms in processing lung CT images. Compared with the existing research, the innovation point of this research is to integrate the improved Inception network with the full convolutional network, solve the problem of low correlation between feature pixels and target pixels by optimizing the structure of the convolutional layer and the pooling layer, and propose the combination of ResNet module and Inception module. To solve the problem of weak information gradient of Inception module. These improvements make the algorithm proposed in this study better than the existing algorithms, such as ResNet and GoogLeNet.

The research is divided into four parts. The first part analyzes the current research status of Inception algorithm and lung CT image classification models. The second part describes Inception and the CT image classification model constructed after fusion with a fully convolutional network. The third part analyzes the performance of the improved Inception algorithm and the improved Inception lung CT image classification model. The last part is a summary of the research.

II. Related Works

Imaging diagnosis of pneumonia remains a major medical challenge. Pneumonia is closely related to human life safety. The prevention and diagnosis of pneumonia are extremely important. The emergence and application of artificial intelligence provide technical support for the treatment of lung diseases. Kocks et al. proposed a machine learning algorithm tool for the diagnosis of asthma, COPD, and overlapping conditions, and compared it with primary care physicians and pulmonary specialists. The results show that the algorithm tool has excellent diagnostic performance and provides strong support for the identification of pulmonary diseases [6]. To solve the problem of inaccurate prediction of lung diseases, Annamalai's team proposed a convolutional neural network based on auction optimization algorithm. The results show that this method is superior to traditional methods in predicting performance and provides a new idea for the early diagnosis

of lung diseases [7]. Wobma et al. developed a screening algorithm for lung disease screening in patients with systemic juvenile idiopathic arthritis. The results show that the algorithm can effectively identify lung diseases in patients and improve the accuracy and efficiency of screening [8]. Zhang et al. reviewed machine learning applications for chronic obstructive pulmonary disease. The results show that machine learning has a broad application prospect in the diagnosis, treatment and management of this disease, providing strong support for clinical decision-making [9]. In order to improve the accuracy of inference for chronic obstructive pulmonary disease, Cosentino et al. proposed a deep learning model based on the original respiratory map. The results show that this method can identify new genetic loci and improve the risk model, providing a new way for the precision medicine of diseases [10].

With the improvement of medical data resources and the development of auxiliary diagnosis and treatment methods for lung diseases, lung CT medical imaging technology plays a crucial role in clinical medicine and lung disease screening. An accurate image classification model can improve the utilization rate of medical resources. To address the shortcomings of training data and its inter class similarity and variability in chest CT scans, Wang et al. proposed a two-stage deep learning strategy. It is used for invasive classification of solid nodules in chest CT images. The results show that the accuracy of the proposed binary classifier is 83.4% ± 1.4, and the AUC is 91.3% ± 2 [11]. To explore the imaging characteristics of congenital tracheal stenosis combined with left pulmonary artery suspension surgery, Fukushima team proposed a qualitative variable model based on clinical features and computed tomography imaging. After testing the model, it was found that 38 cases (86%) of bilateral lung patients has left pulmonary artery suspension surgery [12]. To determine the overall positive rate of pulmonary arteriography and the factors that affect the results of pulmonary arteriography, Aggarwal et al. conducted experiments using methods such as logistic regression analysis. The results showed that the pulmonary angiography rate in the study was 10.9%, higher than the study results with a positive rate of<2% [13]. To explore the role of preoperative CT examination results in predicting acute postoperative exacerbation in patients with gestational pneumonia and cancer, Ozawa et al. used Fisher exact test and Mann-Whitney U test. The test results show that common patterns of gestational pneumonia show 58% acute exacerbation and 73% no acute exacerbation [14]. To accurately estimate the volume of pulmonary nodules and reduce the workload of doctors, Lee proposed an intelligent method for automatically detecting and segmenting pulmonary nodules from computer tomography. The results show that the average and standard deviation of the segmentation overlap measure are 0.81 and 0.08, respectively [15].

In summary, artificial intelligence is applied to all aspects of society. In the era of data and information, people are increasingly reliant on deep neural networks. The improved concept network model is widely used in various fields such as biological component detection, adaptive learning models, and biological gene networks. With the increasing attention paid to lung health, more experts and scholars have conducted experimental research to find the optimal lung CT image classification model. However, lung CT image classification models based on improved Inception have been rarely studied. To make up for this deficiency, a lung CT image classification algorithm model based on an improved Inception network is constructed, thereby promoting the organic integration of artificial intelligence and medicine.

III. Construction of an Improved Inceptional Lung CT Image Classification Model

Nowadays, the application of artificial intelligence in medical images is also gradually increasing. Among them, convolutional neural networks have made significant contributions to medical image processing [16]. The algorithm network based on Inception can obtain deeper level data from medical images, improving the accuracy of image judgment.

A. Construction of a Lung CT Image Classification Model based on Traditional Inception

To integrate the Inception module, convolutional layers and pooling layers are fused together. The pooling layer mainly involves repairing feature images to enhance the correlation between feature pixels and target pixels [17]. This method can effectively save the central and surrounding features of the feature image, avoiding problems such as traditional maximum or average pools ignoring partial feature information [18]. The traditional maximum pooling operation is shown in Fig. 1.

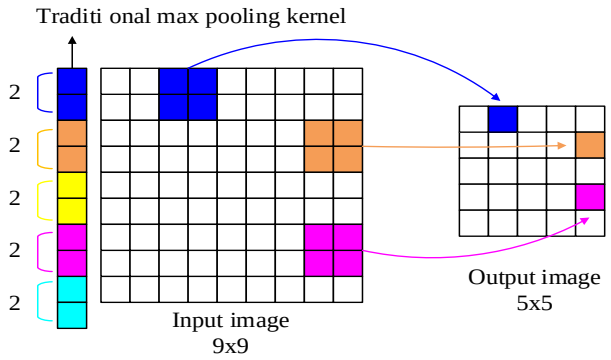


Fig. 1. Traditional maximum pooling operation.

Fig. 1 shows the traditional maximum pooling operation. The left side of Fig. 1 shows the input image. The right side shows the output image. The input image usually has a 9x9 pool core. The pool core size is uniform and fixed in distribution. The output image usually has a 5x5 pool core. The distribution of the output pool cores is uniform. The purpose of pooling layers is to reduce the dimensionality of convolutional layers and reduce the parameters during operation. Afterwards, residual modules and skip connections are used to increase the number of network layers. The calculation of the residual network is shown in Eq. (1).

$$F(x_1) = w \cdot x_1 + b \tag{1}$$

In Eq. (1), x_1 is the input value. w is the weight. b is offset. After the learning signal of the neuron is transmitted to other neurons, the activation function is used to calculate the

sum of the two branch layers. The input data is subjected to data transformation. The activation function is shown in Eq. (2).

$$y_1 = R(F) + h(x_1) \tag{2}$$

In Eq. (2), y_1 represents the input data transformation. y_1 is the sum of two branch layers. After data input and activation, perform residual module output, and the residual module output formula is shown in Eq. (3). It can effectively accelerate training speed and improve the accuracy of parameter adjustment.

$$x_{l+1} = R(y_1) \tag{3}$$

In Eq. (3), x_{l+1} is the output of the residual module. To prevent problems such as gradient dispersion, activation function is used. To enhance the propagation speed of the function, ReLU activation function is used to reduce gradient attenuation caused by deep convolution. The formula for ReLU is shown in Eq. (4).

$$R(x) = \max(0, x) \tag{4}$$

According to Eq. (4), when $x > 0$, $R(x) = x$, the derivative is 1. When $x < 0$, $R(x) = 0$, the derivative is 0. Traditional pooling cannot automatically adapt to changes in the specific location of pooling. Therefore, the Inception module adopts central pooling to preserve the central features of the patch. The central pooling operation process is shown in Fig. 2.

Fig. 2 shows the operation process of central pooling. A large pool core and a small pool core are designed in the central pool structure. The image center selects a small pool core. The image edge adopts a large pool core. This structure can eliminate many edge features. The operation steps first define the pooling area. Pooling operations are typically performed within a certain area of input data. The defined indicators are the size and shape of the area. Rectangles and circles are usually common shape features. Secondly, the total pooling value is calculated. Based on the specific pooling type, a pooling value is calculated within the pooling area. Common pooling types include maximum pooling, average pooling, and L2 pooling. Next, the summary value is output, which represents the information within the pooled area. Finally, the pooling window is moved. The pooling window is moved to the next position of input data. Afterwards, the above steps are repeated until they cover the entire input data. After the operation is completed, in order to avoid information loss during repeated down sampling and up sampling, a fusion multi-convolutional layer feature method is adopted to increase the flow ability of the data image. The specific operation is shown in Fig. 3.

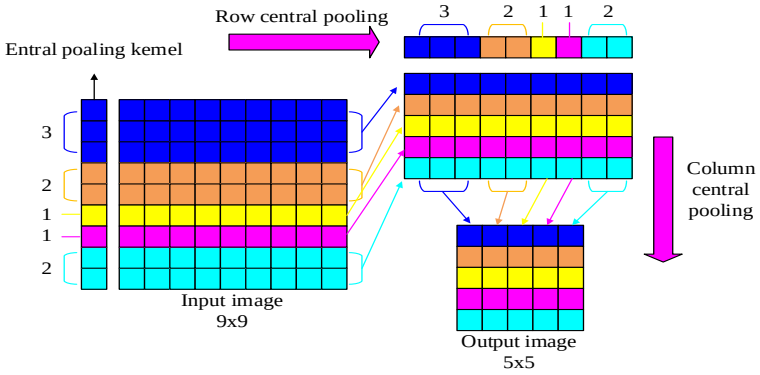


Fig. 2. Central pooling operation.

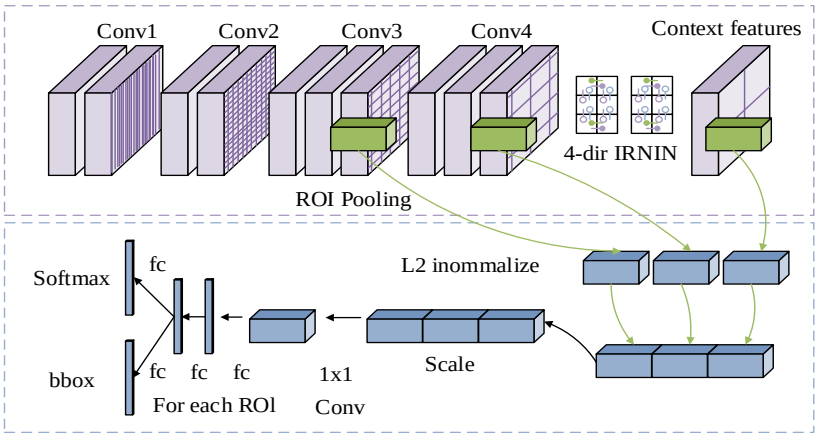


Fig. 3. Inside-outside network.

Fig. 3 shows the inner and outer network structure of the shallow convolutional layer in the Inception module using feature maps. Fig. 3 fuses multiple convolutional layers. The upper and lower features in the image are extracted and normalized in the ROI pool. After the Inception network collects features from the input image in parallel, multiple such Inception modules are first connected in series. Afterwards, all output results are concatenated into feature maps. Different sizes of convolutional kernels are beneficial for diversified feature extraction. The method of collecting signal features in parallel can effectively increase the width of the network and the adaptability to scale. This can effectively reduce information loss caused during the data sampling process, which is more conducive to machine learning.

B. *Construction of an Improved Inception Lung CT Image Classification Model*

The Inception module has four generations, namely the original version, Inception-v1, Inception-v2/v3, and Inception-v4. During the iteration process, the performance of Inception is gradually improved [19]. The traditional Inception model requires a large amount of computation. The information gradient of network backpropagation is weak. On this basis, the Inception model is improved [20-21]. Because ResNet can introduce residual connections, the network can directly transfer the features of the front layer to the back layer, thus effectively alleviating the problem of gradient disappearance. Moreover, it can enhance the feature extraction ability of the network by transmitting feature information directly. In the classification of lung CT images, the accuracy and diversity of feature extraction are very important to improve the classification performance. ResNet's structural design allows the network to more efficiently capture features at different levels in the image, which is very helpful for distinguishing subtle lung lesions. Therefore, the improved method proposed in this study is to integrate Inception with ResNet for better image classification and target detection. The improvement method is to fuse Inception with ResNet for better classification, object detection, and image segmentation. Then the corresponding information of Inception and ResNet module is overlaid. The specific algorithm is shown in Eq. (5).

$$M = \sigma(C_2\theta(C_1 \cdot x)) \tag{5}$$

In equation (5), M represents the output function. σ represents a nonlinear layer. C_2 is the scaling parameter. θ represents the activation function value. x represents the output result. In classification calculations, the loss function is used to standardize two probability distributions a and b. Then b is used to describe a. The cross entropy formula is shown in Eq. (6).

$$H(a,b) = -\sum a(x)\log b(x) \tag{6}$$

$H(a,b)$ represents the cross entropy value. $a(x)$ represents the probability distribution. $\log b(x)$ represents the probability distribution of b. Cross entropy, as a measure of the similarity between two probability distributions, is widely used in machine learning. It is mainly used to measure the difference between the predicted output of the model and the actual output. In Eq. (7), the cross entropy satisfies the probability distribution function.

$$\forall xa(X-x) \in [0,1] and \sum a(X-x) = 1 \tag{7}$$

In Eq. (7), X is the true value predicted by the model. b is the output value predicted by the model. $a(X-x)$ is the distribution function. Eq. (7) indicates that the probability of events occurring is between 1, and 0. The sum of probabilities is always 1. The improved Inception module is mainly divided into six steps to complete. The schematic diagram of the model is shown in Fig. 4.

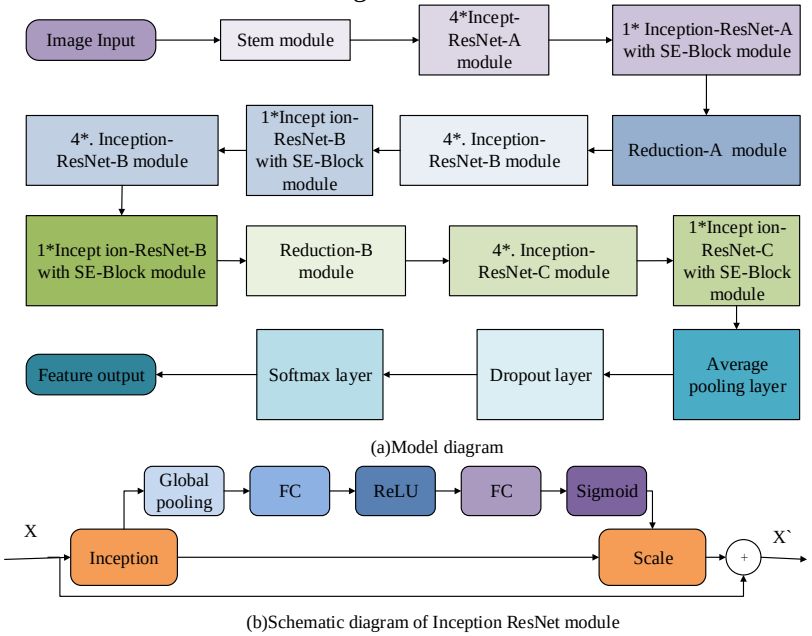


Fig. 4. Module diagram

Fig. 4(a) is a schematic diagram of the convolutional neural network architecture. Fig. 4(b) is a schematic diagram of the Inception ResNet module. Inception and ResNet network structures are combined, and sections such as Stem, SE-Block, and Reduction are added to them. To increase the running speed and deepen the network depth, the convolution size is changed in the model. The residual structures are introduced to solve the overfitting problem. Then, feature weights are added and activation functions are replaced. Finally, the number of classifications is determined. To achieve high network convergence speed in model training, all convolutional layers and inner product activation layers are used in the original U-Net network structure. The normalization formula used is shown in Eq. (8).

$$H_{ab}^{c} = \frac{H_{ab}^{c} - \mu_{ab}}{\sigma_{ab}} \tag{8}$$

In Eq. (8), σ_{ab} represents the output when a neurons in the b layer are trained to c data. μ_{ab} is the average output of a neurons in layer b. σ_{ab} represents the standard deviation of a neurons in layer b. The average output of neurons is shown in Eq. (9).

$$\mu_{ab} = \frac{1}{d}\sum_{c=1}^{d} H_{ab}^{c} \tag{9}$$

In Eq. (9), μ_{ab} is the average output of a neurons in the b layer. σ_{ab} represents the output when a neurons in layer b are trained to c data. $\frac{1}{d}$ represents the mean after summing σ_{ab}. After obtaining the mean formula of the neuron, the standard deviation of the neuron output is calculated. The standard deviation formula is shown in Eq. (10).

$$\sigma_{ab} = \sqrt{\sigma + \frac{1}{d}\sum_{c=1}^{d}(H_{ab}^{c} - \mu_{ab})^{2}} \tag{10}$$

In Eq. (10), σ_{ab} is the standard deviation of a neurons in layer b. μ_{ab} is the average output of a neurons in layer b. The values of σ and d are constants. The specification layer is placed before the activation function. The activation function of the specification layer is normalized in batches, as shown in Eq. (11).

$$y = f(BN(\sum_{a} W_a X_a)) \tag{11}$$

y represents the normalization function. $W_a X_a$ represents the connection degree between input neurons and output neurons. The bias parameter b is normalized by the system after passing through the specification layer. The main task of the BN layer is to normalize the input data values of neurons. The output values are unified within the interval of [0,1] to improve model accuracy. After building the model, the system development environment is built. In the pneumonia classification system, the B/S architecture is selected. The specific system architecture diagram is shown in Fig. 5.

Fig. 5 shows the UML sequence diagram of the pneumonia classification and prediction module. The system page includes two aspects, namely "Select File" and "Forecast". Users can click and select the corresponding test chart in the "Select File" section. The situation where there are a large number of categories in lung CT images that cannot provide classification entry points for each category can be avoided. Afterwards, the "Forecast" button is clicked to perform prediction analysis. When the user enters the platform homepage, the predicted results can be obtained. The controller calls the query interface of the experimental model to obtain CT images. The view layer posts the transmitted information on the page. Finally, the predicted results of pneumonia CT images are obtained. The corresponding probability sizes of the results are displayed in the interface. The graph cutting algorithm is used in the image, which locks and marks the target of the image. The formula is shown in Eq. (12).

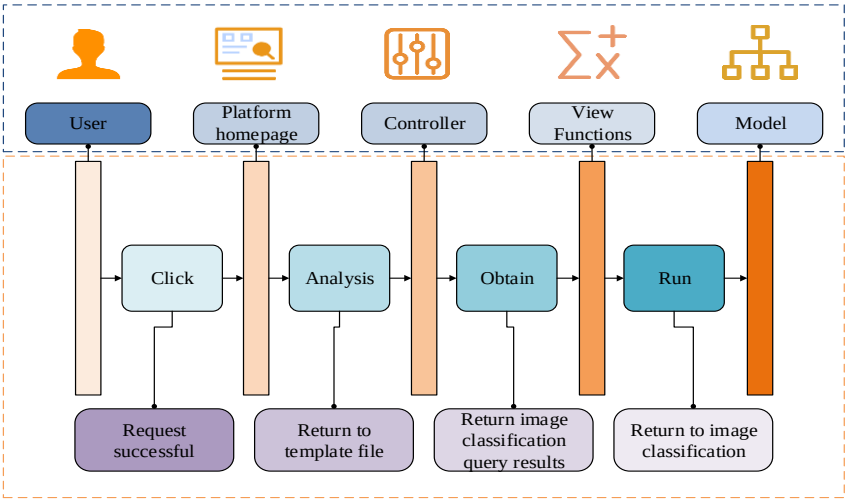


Fig. 5. UML sequence diagram of pneumonia classification and prediction module.

$$E(A) = \lambda R(A) + B(A) \tag{12}$$

In Eq. (12), A represents pixels. B represents the background pixel. $E(A)$ is the energy function. λ is the influence coefficient of the regional energy function on the total energy function. The energy function of the region term is shown in Eq. (13).

$$R(A) = \sum R_P(A_P) \tag{13}$$

In Eq. (13), $R(A)$ represents the regional energy function. R_P represents the probability of the target area. A_P represents the division of pixels. Next, the pixels are labeled as background points and front attractions. The formal expression of $R_P(A_P)$ is shown in Eq. (14).

$$R_p(O) = -\ln \Pr(I_P / O) \tag{14}$$

In Eq. (14), $\Pr$ is the target probability. R_p is the probability result. O represents the pixel of the image. After the "prediction" button is clicked, corresponding prediction analysis is performed, and then the experimental model is called using the controller. After the calculation is completed, the prediction results of the pneumonia CT image are displayed in the front interface. In summary, traditional Inception networks have low classification efficiency. It is not conducive to saving medical resources. Based on this, the traditional Inception network is improved. The improved U-Net full convolutional network image semantic segmentation is integrated into the model to improve the convergence speed of the network. This solves the segmentation speed, laying the foundation for building the optimal lung CT image classification algorithm model.

IV. Evaluation of Lung CT Image Classification Algorithm Based on Improved Inception Network

The improved Inception network lung CT image classification algorithm model can effectively enhance the accuracy of image classification. Compared to other models, the classification performance and fitting degree are superior. It requires fewer parameters. Building a system model framework on the improved U-Net is beneficial for enhancing the sensitivity of image segmentation and making the lung structure clearer.

A. Parameters Design and Evaluation of Improved Inception Network

To avoid inconsistent size and singularity in the lung CT dataset, a pneumonia CT database is constructed for abnormal pneumonia, viral pneumonia, and normal images. The specific content is shown in Table I.

TABLE I. Database

Classification	Normal CT images	Normal CT images	Abnormal CT images
Total	2300	2200	2300
Training set	1500	1300	1600
Validation set	500	500	500
Test set	600	600	600

Table I shows the pneumonia CT database constructed during this training. A total of 6800 pneumonia CT images are selected, including 4400 for training, 1500 for validation, and 1800 for testing. After preprocessing the data, 4400CT images are selected for training. The classification results of ResNet, GoogLeNct, and Inception ResNet are compared. The results are shown in Fig. 6.

Fig. 6 shows the recall rate of the classification results of the three algorithms in CP. The recall rates of ResNet, GoogLeNct, and Inception-ResNet in CP are 0.812, 0.701, and 0.953, respectively. In NCP, the recall results of the three algorithms are 0.754, 0.635, and 0.981, respectively. In the recall rate results of Normal, the Inception-ResNet algorithm is 0.121 and 0.213 higher than ResNet and GoogLeNct algorithms, respectively. From this, it can be seen that the Inception ResNet algorithm in this research has the highest recall rate in CP, NCP, and Normal. Therefore, it has the most superiority. Then, the error and specificity of the three algorithms are compared and analyzed. The results are shown in Fig. 7.

Fig. 7 shows the comparison results of error rates and specificity among the three calculations. Fig. 7(a) shows the comparison results of error rates among the three algorithms. In CP, the training error of Inception-ResNet is 0.013%, the error rate of GoogLeNct is 0.034%, and the error rate of ResNet is 0.023%. In NCP, the error rates of Inception-ResNet are 0.012% and 0.021% lower than those of ResNet and GoogLeNct algorithms. In Normal, the error rates of Inception-ResNet are 0.014% and 0.022% lower than those of ResNet and GoogLeNct algorithms. From this, the Inception-ResNet algorithm in this research has a smaller error rate compared to the other two algorithms. In Fig. 7(b), the Inception-ResNet in CP is 90.1%, which is about 31% and 42% higher than ResNet and GoogLeNct. In NCP, it is about 29.8% and 36.2% higher than ResNet and GoogLeNct. In Normal, it is about 31.2% and 34.3% higher than ResNet and GoogLeNct. From this, the Inception-ResNet algorithm has better classification performance. To further analyze the segmentation performance of Inception-ResNet algorithm in pneumonia CT images, CNN, FCN, GraphCut, region growth value, adaptive threshold and other algorithms are selected for comparison. The results are shown in Fig. 8.

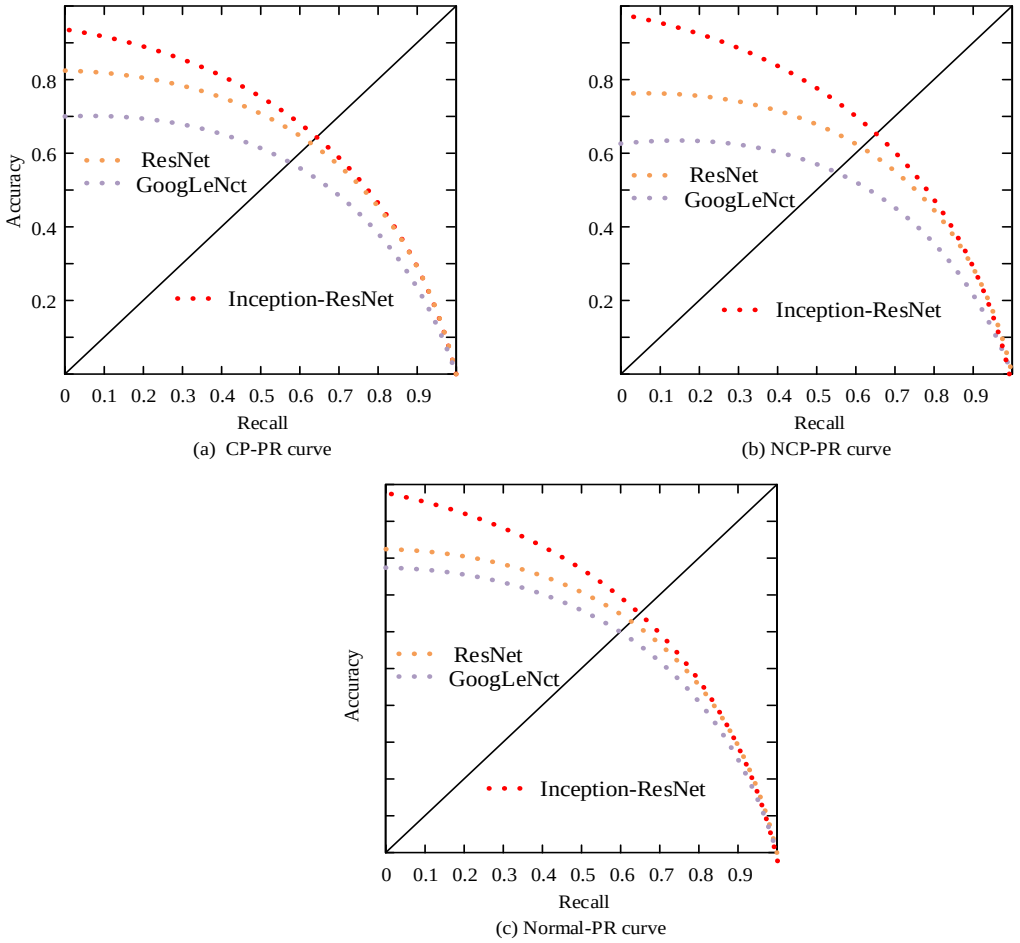


Fig. 6. Recall three algorithms.

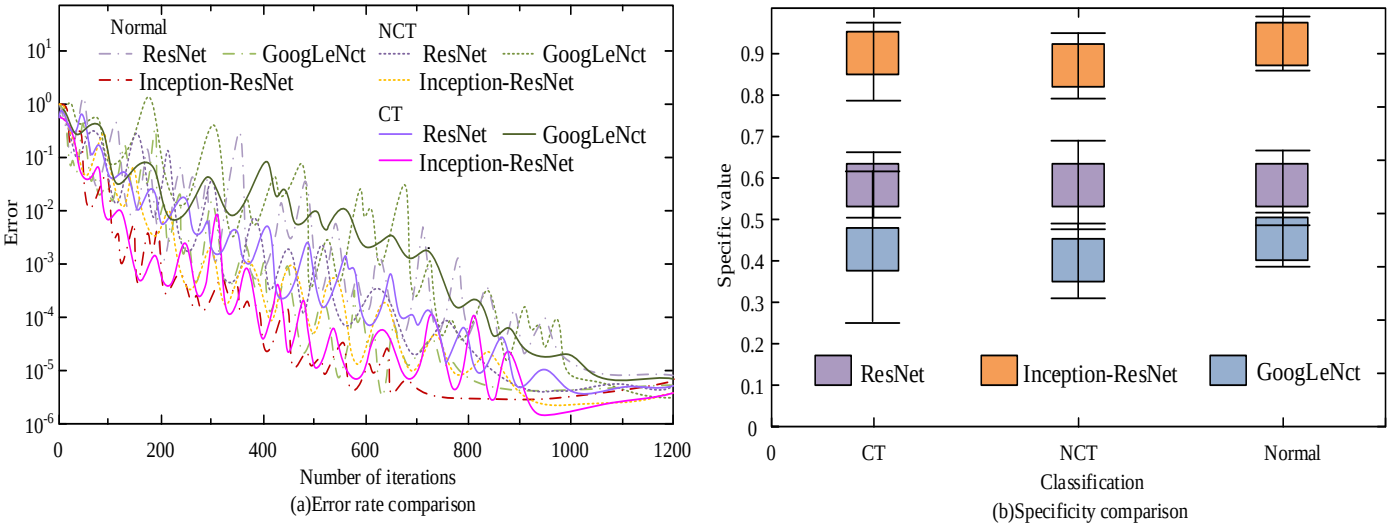


Fig. 7. Comparison of error and specificity.

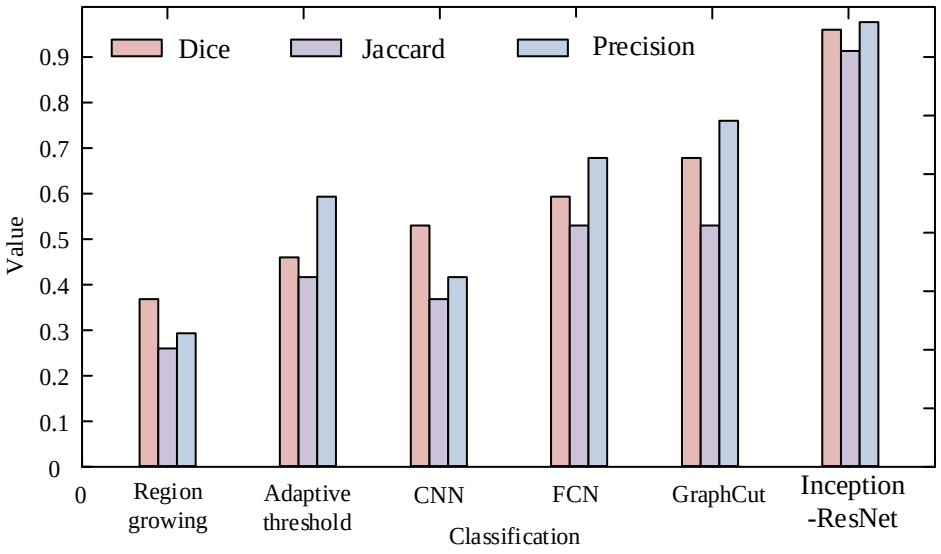


Fig. 8. Algorithm comparison histogram

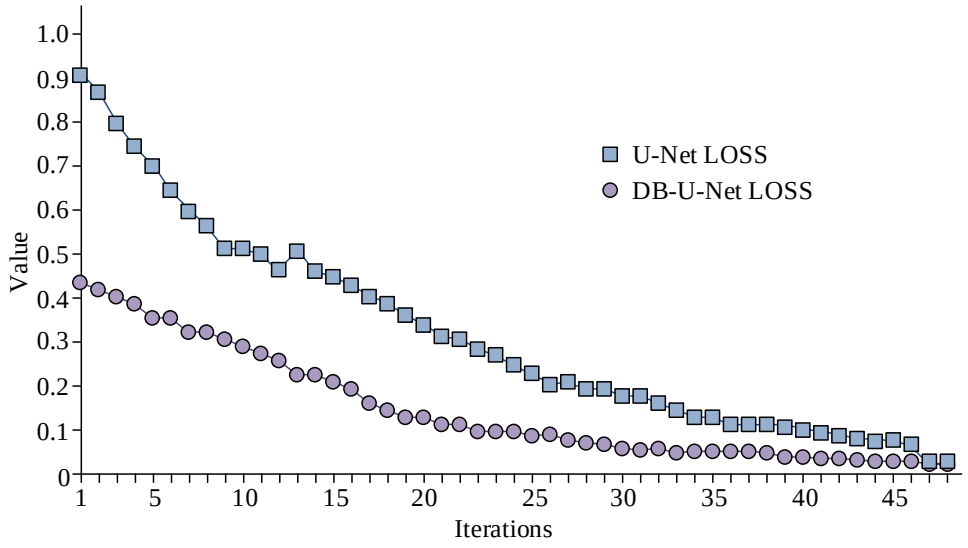
Fig. 8 shows the comparison between the Inception-ResNet algorithm and other algorithms. Compared to other algorithms, Inception-ResNet algorithm has a leading advantage in segmentation accuracy. In the comparison of Dice coefficient and Jaccard coefficient, the Inception-ResNet algorithm is about 0.25 and 0.29 higher than other algorithms, respectively. The Dice coefficient, Jaccard coefficient, and accuracy are the lowest in regional growth. The Dice coefficient, Jaccard coefficient, and accuracy values of the CNN algorithm are 0.51, 0.38, and 0.40, respectively. The Dice coefficient, Jaccard coefficient, and accuracy values of the FCN algorithm are 0.6, 0.54, and 0.68, respectively. The Dice coefficient, Jaccard coefficient, and accuracy values of the GraphCut algorithm are 0.69, 0.51, and 0.77, respectively. The Dice coefficient, Jaccard coefficient, and accuracy values of the Inception-ResNet algorithm are 0.93, 0.91, and 0.96, respectively, all exceeding 0.9. The Inception-ResNet algorithm converges faster, which is beneficial for the segmentation of lung images.

B. Performance Analysis of Improved Inception Network Model

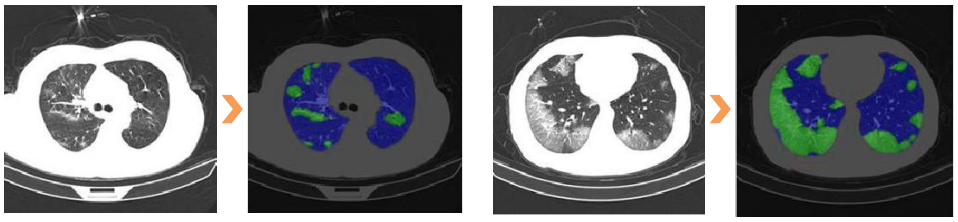
To evaluate the segmentation effect of the improved Inception network model on pneumonia CT images, segmentation experiments are conducted on the model based on image segmentation techniques. During the experiment, the loss function curve was used to represent the iterated values, and the original U-Net is compared with the improved U-Net data. The results are shown in Fig. 9(a). Afterwards, segmentation is performed on different positions of the lungs to segment the main body of the lung CT image. The final result is shown in Fig. 9(b).

Fig. 9(a) shows the loss function of the network model. In model comparison, the convergence speed of the original network model is slower than the DB-U-Net Loss model proposed in the research. When the DB-U-Net Loss model is less than 0.1, the loss function converges rapidly and reaches a minimum value around 0.18. When the number of iterations is 20, it approaches 0. The original model approaches 0 at 45 iterations. The convergence speed is relatively slow. This indicates that the improved model has better network performance. Fig. 9(b) shows the segmentation results of pneumonia images. The improved model can accurately segment lesions with pneumonia features, which has a high degree of segmentation granularity. When the segmentation model reaches the optimal level, collect the operational status of pneumonia CT images within 48 hours. The average response time, fault condition, and waiting time for the slowest response of the system are used for data comparison. The analysis results of the model performance are shown in Fig. 10.

The white screen time and first screen time in Fig. 10 reflect the response speed of the system when processing CT images of pneumonia. The white screen time refers to the time from the user's request to the system's content display, and the first screen time refers to the time from the system's content display to the completion of the main page content loading. The length of these two times directly affects the user experience, and the shorter the time, the faster the system response and the better the user experience. Fig. 10 shows the performance analysis of the model system proposed in this study. As shown in the figure, the average first screen time in the pneumonia CT image classification system is 1.01s. The average white screen time is 0.81s. The average full screen time is 1.36s. The full screen time is the longest, approximately 0.35s and 0.55s higher than the average white screen time and average first screen time, respectively. As the iteration increases, the white screen time is the most stable. The overall screen time fluctuates significantly. The first screen time fluctuates greatly when the number of iterations is 2-30. This functional test shows that the performance of each functional module and page is normal. There is no malfunction in the reaction sensitivity, and the reaction force is sensitive. Finally, the analysis is conducted using a confusion matrix approach. ResNet and GooLeNet models are compared with the Inception-ResNet model for image classification accuracy. The results are shown in Fig. 11.



(a)Iterative Network Model Loss Function



(b)Pneumonia image segmentation

Fig. 9. Loss function and image segmentation.

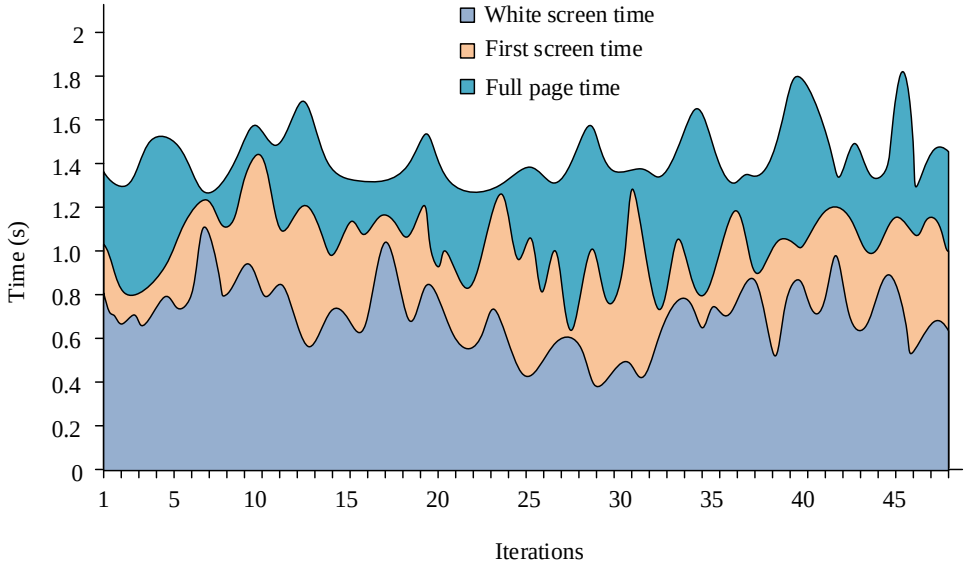


Fig. 10. Model system performance analysis.

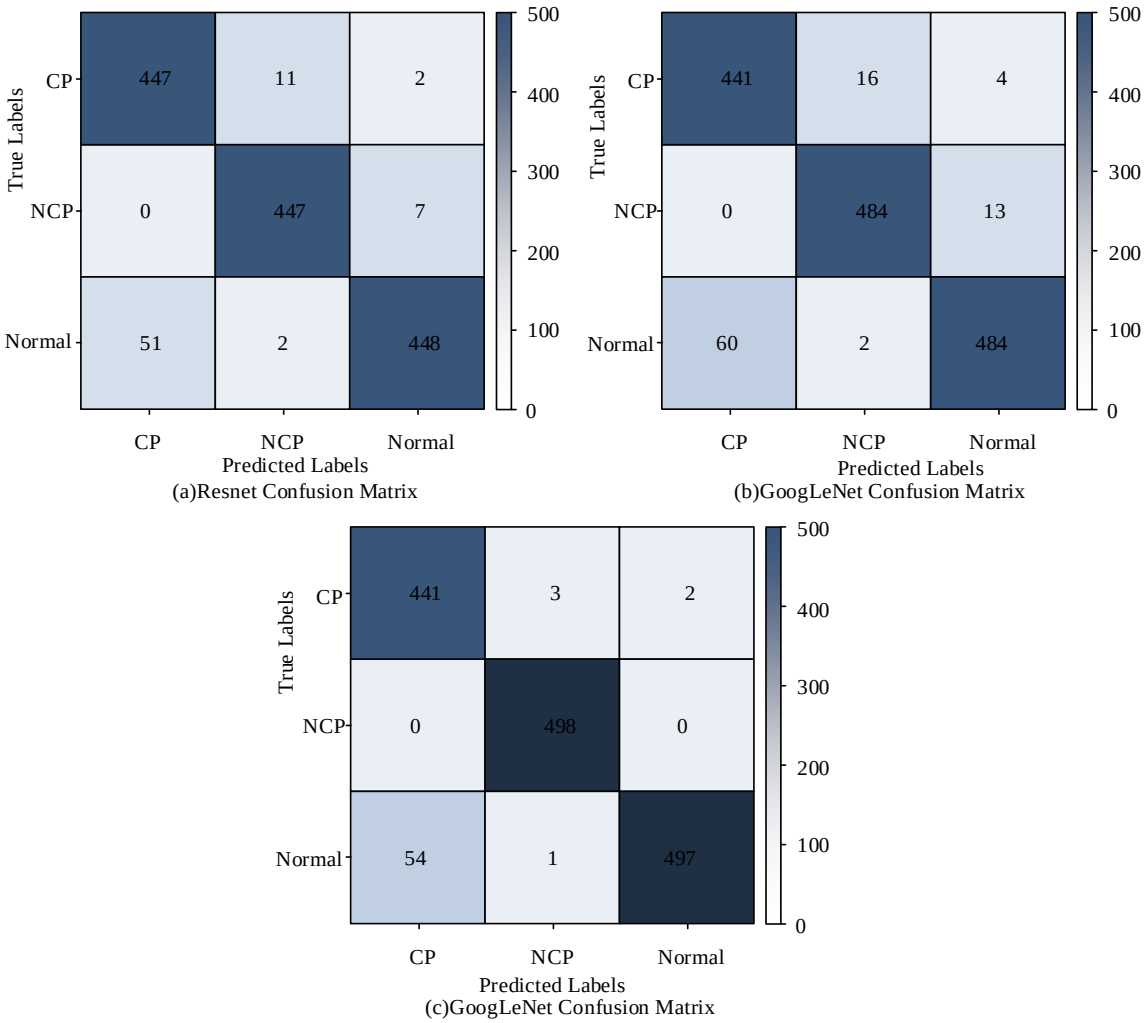


Fig. 11. Comparative analysis of confusion matrix.

Fig. 11(a) shows the confusion matrix of the ResNet model. The accuracy rate of non-severe viral classification is 89.7%. The accuracy rate in classifying abnormal pneumonia images reaches 96.3%. Fig. 11(b) shows the confusion matrix of the Google Net model. It can be seen that the classification accuracy of viral pneumonia using the Google Net model is 88.1%. The accuracy rate of classifying abnormal pneumonia is 88.7%. The classification accuracy of viral pneumonia using the Inception-ResNet model is 98.1%. The accuracy rate of classifying abnormal pneumonia is 98.7%. This method has higher classification accuracy than the other two models for pneumonia images. Compared to other convolutional neural network models, the Inception-ResNet model has higher accuracy, faster processing speed for image classification, and more mature image segmentation technology. In the performance analysis of the system model, the Inception-ResNet model exhibits a zero error rate and extremely high sensitivity. When compared with algorithms such as ResNet and GoogLeNct, the Inception-ResNet algorithm has a lower error rate and higher accuracy and specificity.

V. Discussion

In this study, a lung CT image classification algorithm model based on improved Inception network was proposed, and its performance was verified by experiments. The experimental results show that in CP, the recall rate of Inception-ResNet is 0.953, while that of ResNet and GoogLeNet are 0.812 and 0.701, respectively. In Normal, the recall rate of Inception-ResNet was also about 0.121 and 0.213 higher than the other two algorithms. This result is similar to that of Nigudgi et al. [22]. The results show that the epigenetic-RESnet algorithm has higher accuracy in identifying various types of lung CT images, which is mainly due to the advantages of integrating Inception and ResNet modules to improve the diversity and accuracy of feature extraction. In addition, from the comparison of error rate and specificity, the error rate of the initiation-Resnet algorithm in CP, NCP and Normal is lower than that of the ResNet and GoogLeNet algorithms, and the specificity is significantly higher than the other two algorithms. Especially in CP, the Inception-ResNet error rate is only 0.013%, while the specificity is as high as 90.1%, which is about 31% and 42% higher than ResNet and GoogLeNet, respectively. This

suggests that the Inception-ResNet algorithm performs better in reducing misjudgments and improving classification specificity, which helps to improve the accuracy of lung disease diagnosis.

In addition, the study is compared with CNN, FCN and GraphCut classification algorithms. The results show that the Inception-ResNet algorithm is superior to other algorithms in Dice coefficient, Jaccard coefficient and precision, especially in segmentation accuracy. This result coincides with the conclusion obtained by Sethy team [23]. This further proves the effectiveness of the Inception-ResNet algorithm in lung CT image classification. In terms of model performance analysis, the loss function of the improved Inception model rapidly converges during training and reaches the minimum value within a few iterations, which indicates that the model has high learning efficiency and stability. At the same time, in the image segmentation experiment, the improved U-Net network can segment the lesions with pneumonia characteristics more accurately, which helps doctors to judge the disease more accurately. This study further proves that the Inception-ResNet model can effectively improve the accuracy and efficiency of pulmonary CT image classification by integrating Inception and ResNet modules. In the future, the combination of more medical data and 3D imaging technology is expected to further improve the generalization ability and diagnostic accuracy of the model, and provide more powerful technical support for clinical lung disease diagnosis.

VI. Conclusion

To solve the complex images, small sample data, and similar symptoms between diseases in traditional lung CT medical image processing, the improved Inception algorithm is integrated with the improved U-Net fully convolutional network. The aim is to construct a lung CT image classification model based on an improved Inception network to improve image processing efficiency and lung disease diagnosis speed. In training, the Inception algorithm is compared and analyzed with other algorithms in terms of performance. The results show that the Inception-ResNet algorithm is approximately 0.121 and 0.213 higher than ResNet and GoogLeNct algorithms, respectively. Among CP, NCP, and Normal, this method has the highest recall rate. In the specificity comparison, the algorithm proposed in this research is 90.1% in CP. In NCP, it is about 29.8% and 36.2% higher than ResNet and GoogLeNct. In Normal, it is about 31.2% and 34.3% higher than ResNet and GoogLeNct. In the comparison of model performance, when it is less than 0.1, the improved Inception model's loss function converges rapidly and reaches a minimum value of around 0.18. When the number of iterations is 20, it approaches 0. The classification accuracy of viral pneumonia using the improved Inception model is 98.1%. The classification accuracy of abnormal pneumonia is 98.7%. In summary, the improved Inception network lung CT image classification model can accurately and efficiently process lung CT image classification, which has high practicality and feasibility. Although this research has achieved some results, the future work still needs to be deepened. On the one hand, it is necessary to continue to optimize the algorithm model and explore higher-dimensional feature extraction methods in the future to further improve the classification accuracy and efficiency. On the other hand, given that the current model is based on 2D images and cannot fully reflect the 3D structure of the lung, future research will focus on applying deep learning technology to the field of 3D imaging in order to achieve more accurate diagnosis of lung diseases. In addition, the dataset needs to be expanded to cover more types of lung diseases in the future to enhance the generalization ability and practicality of the model.

Fundings

The research is supported by: The Scientific Research Project of Department of Education of Hunan Province, Research on brain network based on multi-modal imaging, (No.22C0540).

References

[1] Abdulla A A. Efficient computer-aided diagnosis technique for leukaemia cancer detection. IET Image Processing, 2020, 14(17):178-186.

[2] Kassis R W L. Advances in Diagnosis of Progressive Pulmonary and Disseminated Coccidioidomycosis. Clinical infectious diseases, 2021, 72(6):62-89.

[3] Wang B, Chen K, Tian X, Yang Y, Zhang X. An Effective Deep Network for Automatic Segmentation of Complex Lung Tumors in CT Images. Medical Physics, 2021(3):334-357.

[4] Tang W, Yang Q, Hu X, Yan W. Deep learning-based linear defects detection system for large-scale photovoltaic plants based on an edge-cloud computing infrastructure. Solar Energy, 2022, 231(3):527-535.

[5] Li W C, Yan Q R, Guan Y Q, Yang S T, Peng C, Fang Z. Deep-learning-based single-photon-counting compressive imaging via jointly trained subpixel convolution sampling. Applied optics, 2020(23):59-83.

[6] Kocks J W H, Cao H, Holzhauer B, Kaplan A, FitzGerald J M, Kostikas K, Mastoridis P. Diagnostic performance of a machine learning algorithm (Asthma/Chronic Obstructive Pulmonary Disease [COPD] Differentiation Classification) tool versus primary care physicians and pulmonologists in asthma, COPD, and asthma/COPD overlap. The Journal of Allergy and Clinical Immunology: In Practice, 2023, 11(5): 1463-1474.

[7] Annamalai B, Saravanan P, Varadharajan I. ABOA-CNN: auction-based optimization algorithm with convolutional neural network for pulmonary disease prediction. Neural Computing and Applications, 2023, 35(10): 7463-7474.

[8] Wobma H, Bachrach R, Farrell J, Chang M H, Day-Lewis M, Dedeoglu F, Henderson L A. Development of a screening algorithm for lung disease in systemic juvenile idiopathic arthritis. ACR Open Rheumatology, 2023, 5(10): 556-562.

[9] Zhang B, Wang J, Chen J, Ling Z, Ren Y, Xiong D, Guo L. Machine learning in chronic obstructive pulmonary disease. Chinese Medical Journal, 2023, 136(5): 536-538.

[10] Cosentino J, Behsaz B, Alipanahi B, McCaw Z R, Hill D, Schwantes-An T H, Hormozdiari, F. Inference of chronic obstructive pulmonary disease with deep learning on raw spirograms identifies new genetic loci and improves risk models. Nature Genetics, 2023, 55(5): 787-795.

[11] Wang J, Chen X, Lu H, Zhang L, Pan J, Bao Y. Feature-shared adaptive-boost deep learning for invasiveness classification of pulmonary subsolid nodules in CT images. Medical Physics, 2020, 47(4):113-147.

[12] Fukushima N, Shimojima N, Ishitate M. Clinical and structural aspects of tracheal stenosis and a novel embryological hypothesis of left pulmonary artery sling. Pediatric Pulmonology, 2020, 55(3):231-249.

[13] Aggarwal T, Eskandari A, Priya S. Pulmonary embolism rule out: Positivity and factors affec ting the yield of CT angiography. Postgraduate Medical Journal, 2020, 96(11):1130-1156.

[14] Ozawa Y, Shibamoto Y, Hiroshima M, Nakagawa M, Kitase M. Preoperative CT Findings for Predicting Acute Exacerbation of Interstitial Pneumonia After Lung Cancer Surgery: A Multicenter Case-Control Study. American Journal of Roentgenology, 2021, 217(4);119-134.

[15] Lee C H, Jwo J S. Automatic segmentation for pulmonary nodules in CT images based on multifractal analysis. IET Image Processing, 2020, 14(7):1347-1353.

[16] Fang Y, Luo B, Zhao T, He D, Jiang B, Liu Q. ST-SIGMA: Spatio-temporal semantics and interaction graph aggregation for multi-agent perception and trajectory forecasting. CAAI Transactions on Intelligence Technology, 2022, 7(4):744-757.

[17] Zhou L, Mao Q, Huang X, Zhang, F, Zhang Z. Feature refinement: An expression-specific feature learning and fusion method for micro-expression recognition. Pattern Recognition, 2022, 122(2):134-157.

[18] Tarsitano F, Bruderer C, Schawinski K. Image feature extraction and galaxy classification: a novel and efficient approach with automated machine learning. Monthly Notices of the Royal Astronomical Society, 2022, 511(3):3330-3338.

[19] Torre C, Guerreiro J, Longo P, Raposo J F, Leufkens H, Martins A. Intensive monitoring of adverse drug events associated with the use of new glucose-lowering drugs: results from an inception cohort study in Portugal. Diabetic Medicine, 2020, 37(4):116-179.

[20] Guo X L C. The formation of the steady and unsteady air-entrained vortices in pump sump. International Journal of Multiphase Flow, 2020, 129(1):126-137.

[21] Nimrah S, Saifullah S. Context-Free Word Importance Scores for Attacking Neural Networks. Journal of Computational and Cognitive Engineering, 2022, 1(4): 187-192.

[22] Nigudgi S, Bhyri C. RETRACTED ARTICLE: Lung cancer CT image classification using hybrid-SVM transfer learning approach. Soft Computing, 2023, 27(14): 9845-9859.

[23] Sethy P K, Geetha Devi A, Padhan B, Behera S K, Sreedhar S, Das K. Lung cancer histopathological image classification using wavelets and AlexNet. Journal of X-Ray Science and Technology, 2023, 31(1): 211-221.

Multilevel Characteristic Weighted Fusion Algorithm in Domestic Waste Information Classification

Min Li
School of Computer and Artificial Intelligence, Henan Finance University, Zhengzhou, 451464, China

***Abstracts*—The study of domestic waste image classification holds significant significance for fields like environmental protection and smart city development. To improve the classification efficiency of household waste information, a multi-feature weighted fusion method for household waste image classification is proposed. In this research, deep learning technology was applied to develop a multi-level feature-weighted fusion network model for domestic garbage image classification. The study first analyzed the VGG-16 architecture and created a garbage image dataset for domestic garbage according to the current Shenzhen garbage classification standard. Based on this, a multi-level feature-weighted fusion model for garbage image classification was constructed using VGG-16 as the backbone network. Furthermore, it was combined with the backbone feature extraction network as well as the content-aware and boundary-aware feature extraction networks. The performance of the classification model was tested, and it was found that the highest classification accuracy of the classification model can reach 0.98, and the shortest classification time is only 3s. The multi-level feature-weighted fusion garbage image classification model constructed in this research not only has better classification performance, but also can provide a new processing idea for the urban garbage classification problem.**

Keywords*—Multi-feature; weighted fusion; image; deep learning; waste classification

I. Introduction

With urbanization accelerating, domestic waste disposal presents a growing challenge for city managers [1-2]. Waste classification (WC) can impact resource recovery, environmental protection, as well as the efficiency and quality of city operations. Successful WC entails fast, precise identification of various types of waste, a considerable challenge with the rising quantities of waste. Currently, the utilization of computer vision technology in WC is emerging as a popular research area [3-4]. Traditional garbage image classification methods rely on manually extracting features and using simple classifiers, which are limited in dealing with images that have complex backgrounds, diverse lighting, and different angles. Due to these limitations, the current garbage classification methods have the problems of low accuracy and low efficiency. Recently, convolutional neural networks (CNNs) in deep learning (DL) algorithms have displayed excellent performance for image classification tasks. However, current models encounter issues such as inadequate feature extraction and suboptimal adaptation to limited datasets [5]. In order to address the issue of low-quality garbage image classification, this study employs the VGG-16 model as the primary backbone network. A multi-feature weighted fusion image classification model of domestic waste is constructed by integrating the content and boundary perception model. Against this background, the research objective is to develop a new multi-level feature-weighted fusion network model (MFWFNM) to improve the accuracy and efficiency of domestic garbage image classification (DGIC), and to overcome the limitations of conventional models in handling complicated garbage images. The significance of this study lies in the use of this method to speed up garbage classification, facilitate users to perform garbage classification, create a good living environment, and provide certain ideas and methods for DGIC. The study comprises of six sections, first being the Introduction: Next section presents an analysis and summary of related research from other sources. The third section details the construction of the network model. The model's performance testing is covered in the fourth section, fifth section presents the discussion and the study's overall findings are presented in the last section.

II. Related Work

With the rapid pace of urbanization, the disposal of domestic waste poses a paramount challenge for urban management. Efficient waste management not only facilitates resource recovery and environmental protection, but also increases the productivity of waste disposal. In this regard, automatic classification of garbage images by computer vision technology is an important technical solution to improve the efficiency of waste management. Aiming at the shortcomings of traditional classification models, Ma et al. proposed a WC model that improves the structure of ResNet-50 network. The model was improved in two stages. First, by adding an attention module to the residual block and modifying the downsampling procedure, the input feature screening and information loss reduction were adjusted. Second, the model incorporated horizontal and vertical multi-scale feature fusion techniques to utilize the features more efficiently. The model's classification performance on tiny datasets was significantly enhanced by this restructure. The outcomes revealed that the model outperformed the original ResNet-50 model by 7.62% on the TrashNet dataset, demonstrating higher accuracy and robustness [6]. Wu et al. created a new publicly available benchmark dataset, in addition to the study a classification test was performed on this dataset using deep CNN. The results of the study showed that the created home spam image dataset was able to simulate different lighting, backgrounds, angles, and shapes, and the deep CNN was able to obtain a high classification accuracy (CA) in the data [7]. The rise of email as a means of official and personal communication has made it increasingly difficult to accurately identify and classify spam. Scholar Vivekanandam developed a novel hybrid machine

learning approach to effectively detect spam. This research was significant in safeguarding data security against unauthorized access, while also offering fresh insights and techniques for spam detection [8]. In an effort to address the issues of overfitting, inadequate convergence, and decreased recall and accuracy in conventional image recognition algorithms, Li Y et al. presented a DL-based spam image identification system. The algorithm employed the ReLU activation function to handle the gradient dissipation problem in neural network training, the Adagrad adaptive method to modify the parameters of the deep neural network, and the Dropout strategy to prevent overfitting [9].

In recent years, the use of lightweight networks is pervasive in image segmentation and face recognition due to their advantageous characteristics, including minimal resource consumption and rapid reasoning speed. The utilization of lightweight networks for the purposes of classification and image recognition has emerged as a significant area of research. For use in the facial expression picture identification task, Zou et al. introduced the multi-feature fusion CNN, a lightweight network architecture. According to experimental data, the suggested model performed better on average in terms of recognition accuracy than other compared algorithms [10]. In the face of COVID-19 epidemic, the problem of high dropout rate in catechism classes was becoming more and more prominent. Because of this, Yujiao et al. presented a dropout prediction model that combined support vector machines with feature fusion from behavioral data. The final fused features were created by applying varying weights to various behavioral features in a model that was based on Pearson's principle. The suggested model had a higher accuracy of dropout prediction, according to experimental results [11]. Zhang et al. aimed to develop an intelligent detection model for strabismus based on corneal light reflection photographs. To enhance strabismus detection performance, a multi-feature fusion model is suggested that integrates both ratio and depth features. According to the experimental data, the model successfully increased the accuracy and reliability of strabismus diagnosis, achieving an accuracy of 97.17% in strabismus identification, which was much superior to the single-feature model [12]. A unique multi-feature fusion improved transformer for the creation of image descriptions was proposed by Zhang et al. According to test results on the MSCOCO dataset, the suggested model performed better in picture description than other cutting-edge techniques [13].

In summary, numerous scholars have conducted research on junk image classification and multi-feature fusion algorithms, achieving significant results. The extant research on garbage image classification indicates that, while the results are effective, the CA and efficiency remain at a relatively low level. It is imperative to identify an effective method to enhance the classification of garbage images. Therefore, further exploration of this area appears promising. To enhance CNN's performance in solving the DGIC problem, this study utilizes a multi-feature weighted fusion DGIC model with VGG-16 as the backbone network. This model is combined with content and boundary-aware models to improve the WC model's CA and provide fresh solutions to the real-world WC problem.

III. Domestic Garbage Image Classification Method Based on Multi-Feature Weighted Fusion

To enhance the CA of domestic garbage images (DGI), this research firstly introduces the VGG-16CNN model in DL algorithm and the DGI dataset used in this research. Based on this, the boundary features and contents of the garbage images are fused, and a multilevel feature-weighted fusion algorithm model for DGIC is further proposed.

A. VGG-16 Modeling and Design of Domestic Garbage Images Dataset

The VGG-16 model of CNN is adopted in this study for domestic waste treatment. CNN belongs to the part of artificial intelligence research field, because it has a certain degree of autonomous learning, and can simulate the processing mechanism of the human brain visual nerve, so it is often used in the analysis of visual images [14]. The input layer (IL), convolutional layer (CL), pooling layer (PL), fully connected layer (FCL), and output layer make up the majority of a typical CNN. Fig. 1 depicts its overall structure.

The primary function of the CL in Fig. 1 is to extract features using the convolutional kernel. The retrieved data comes from the input of the IL, and the image features that each layer of the CL can extract are different. CNN backpropagation optimizes the weight coefficients and bias of the convolution kernel, and the size of the convolution kernel controls the complexity of the features that can be recovered from the convolution layer. The primary task of the PL is to select a large number of features to simplify the complexity of the model operation. The PL often comes after the CL. After feature extraction in the CL, a large number of image features will be sent to the PL. PL operation is generally divided into maximum pooling and average pooling, both are a large amount of feature information is divided into a number of small regions. The difference is that the former outputs the maximum value in each small region, while the latter outputs the average value of all feature information in each small region. In addition to the input and output layers, the FCL also plays an important role in CNN. The FCL, which is at the end of the CNN, is responsible for nonlinearly combining all the feature information before feeding it to the output layer. The CL and PL, on the other hand, extract and select image features, respectively.

The VGG-16 model is a popular DL network architecture named after the institution where its development team is located, i.e., the visual geometry group (VGG) at the University of Oxford. VGG-16 refers to the fact that this network structure contains 16 network layers with weights, including convolutional and FCLs. VGG-16 is a unique CNN that is often utilized in domains like image recognition and classification because of its outstanding performance and straightforward architecture. Fig. 2 depicts the VGG-16 network model's structure.

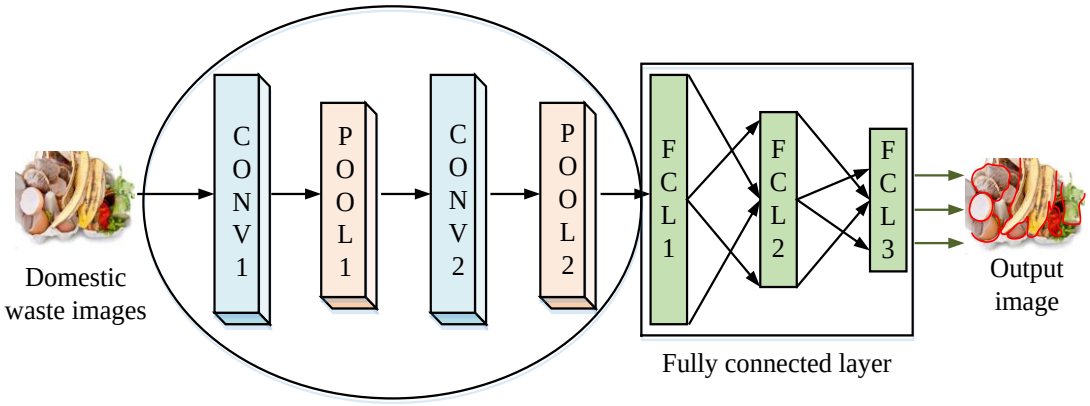


Fig. 1. Basic structure of CNN model.

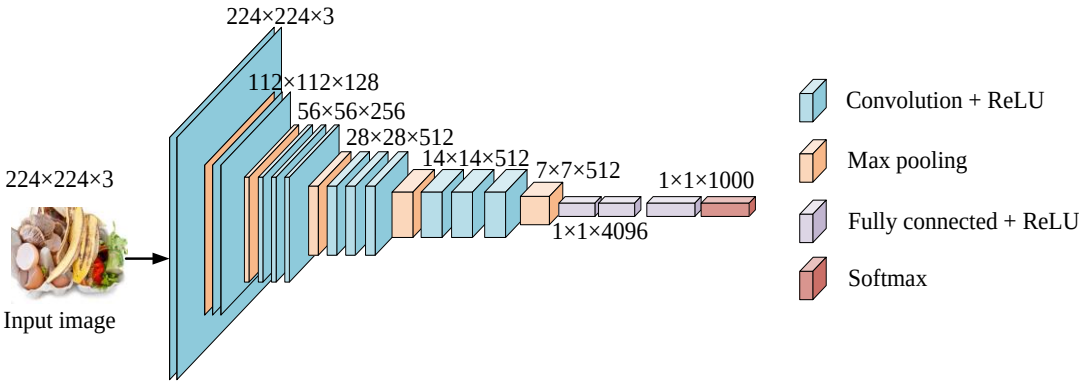


Fig. 2. VGG-16 network model structure diagram.

An FCL, a Softmax layer, an activation function, an IL, a CL, and a PL make up the entire VGG-16 network model shown in Fig. 2. The IL is a fixed-size, 224*224 RGB picture. The whole model has a total of 13 CLs, each CL has a convolutional kernel size of 3*3 with a step size of 1. Using a small convolutional kernel enables the VGG-16 network to have more weight layers, which improves the depth and complexity of the network. Each CL is followed by a ReLU activation function to increase the nonlinearity of the network. The network has a total of five PLs with a filter size of 2x2 and a step size of 2. Behind the convolutional and PLs, there are three FCLs that are connected to a Softmax layer. The Softmax layer outputs the probability values for image classification. Since the VGG-16 network model has a simple operational framework, powerful feature extraction capability and good generalization ability, this study uses the VGG-16 network as the backbone network to build a garbage image classification model.

Due to the wide variety of household garbage categories and different classification standards, the data used in this study is the DGI dataset that Huawei has previously made public in the WC Challenge Cup. This dataset contains more categories of garbage and a richer database than foreign garbage image datasets such as TrashNet and GINI, and contains a total of more than 10,000 DGIs, and the garbage in this dataset is categorized according to the WC standards of Shenzhen City. Additionally, Table I displays its classification standards.

TABLE I. CLASSIFICATION STANDARDS FOR THE DOMESTIC WASTE DATA SET

Garbage classification standards	Contains content
Hazardous waste	Batteries, expired medicines, ointments, etc.
Recyclable trash	Beverage bottles, glasses, cans, pillows, plush items, cardboard boxes, etc.
Kitchen waste	Fruit peels, tea leaves, bones, eggshells, vegetable leaves, leftovers, etc.
Other garbage	Disposable tableware, cigarette butts, toothpicks, plastic products, etc.

As shown in Table I, the waste image dataset used in this study is classified into four major categories: hazardous waste, recyclable waste, food waste, and other waste according to the WC standard of Shenzhen. Due to the large number and variety of waste images in the dataset, it is easy for the existence of similar appearance of different categories of household waste to lead to classification errors. To reduce the experimental errors, 80% of the images in the dataset are randomly and mutually exclusive divided into a training set before the experiment, leaving only 20% of the images as a test set, and the distribution of the dataset does not change during the entire experiment.

B. Design of Multi-Level Feature Weighted Fusion Network Model for Domestic Garbage Image Classification

Classifying DGIs quickly and accurately is a crucial and difficult task in the field of WC and processing nowadays.

Traditional image classification methods often face the problems of poor accuracy and lack of robustness when dealing with spam images with complex backgrounds, diverse lighting conditions and variable shapes. The utilization of multi-feature fusion methods to enhance image classification performance has become a new research trend with the development of DL and computer vision technologies. However, the effective integration of different levels of image features, such as color, shape, and texture, into beneficial information for classification remains a problem to be solved [15]. In this research, the VGG-16 network is used as the backbone network, and four different feature extraction networks are fused to design the DGIC-oriented MFWFNM. Fig. 3 displays the overall categorization model's structure.

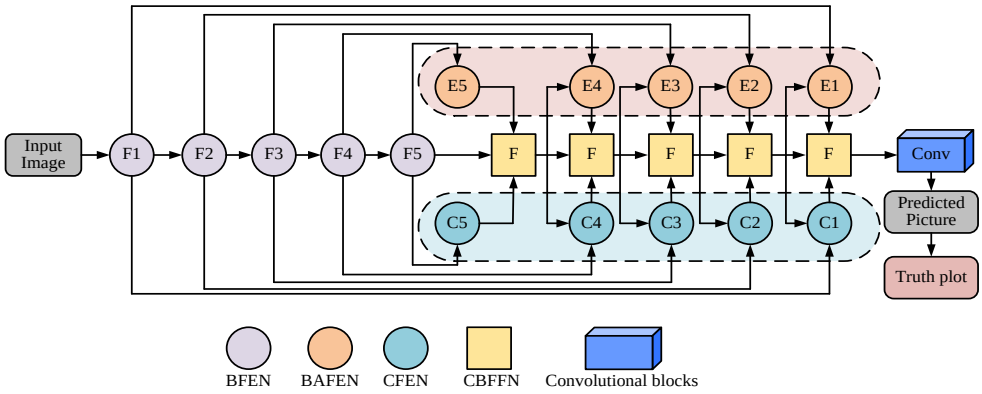


Fig. 3. Structure diagram of multi-level feature weighted fusion network model.

The multilevel feature-weighted fusion network model in Fig. 3 incorporates a total of four feature extraction networks, namely backbone feature extraction network (BFEN), boundary-aware feature extraction network (BAFEN), content-aware feature extraction network (CFEN), and content boundary feature fusion network (CBFFN). The garbage images from the domestic waste dataset are selected and input into the classification model, firstly, the multi-level features are extracted by using BFEN in VGG-16 network, then the effective content is extracted from the main features by using CFEN, and the edge information is extracted from the main features by using BAFEN. Then CBFFN is used at each stage to screen and fuse the features. Lastly, the model's ability to categorize the garbage photos is tested using a bespoke loss function. Training is stopped whenever the model is able to do so. Fig. 4 depicts the precise CFEN structure.

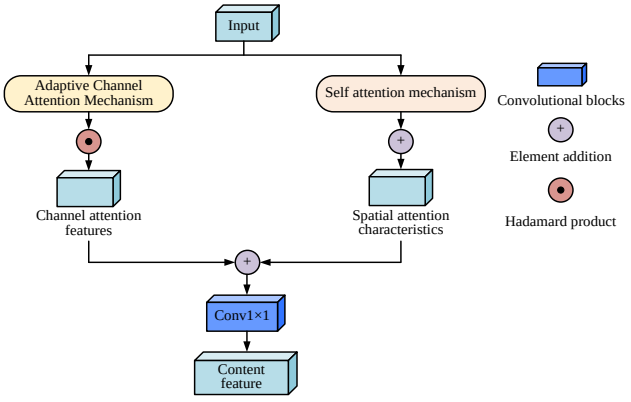


Fig. 4. CFEN structure diagram.

In Fig. 4, the CFEN is mainly composed of the channel attention feature extraction module and the spatial attention feature extraction module. For layer i, the fusion feature S_{i+1} from the previous layer of CBFFN is used as the input of CFEN, and the channel attention mechanism and spatial attention mechanism are utilized to complete the fusion of the features, so as to obtain the new features. Assuming that the feature of the output channel is F_C, the calculation equation of the channel attention feature extraction module is obtained as shown in Eq. (1).

$$F_C = S_{i+1} \text{Å} CA(S_{i+1}) \tag{1}$$

In Eq. (1), CA denotes the adaptive channel attention mechanism. Å denotes the inter-pixel addition operation. Eq. (2) displays the computational equation for the feature extraction module of spatial attention.

$$F_S = S_{i+1} \text{Å} SA(S_{i+1}) \tag{2}$$

In Eq. (2), F_S denotes the output spatial features. SA denotes the adaptive spatial attention mechanism. The extracted channel dimension and spatial dimension features in Eq. (1) and Eq. (2) are fused to obtain the equation of the fused features as shown in Eq. (3).

$$C_i = conv(F_C \text{Å} F_S) \tag{3}$$

In Eq. (3), $conv(\times)$ denotes the convolution operation and the size of the convolution kernel is 1*1. C_i denotes the content fusion feature.

In addition to using the CFEN model to fuse content features, it is also necessary to extract boundary features from garbage images. This study employs the BAFEN network to supplement edge information in order to extract more meaningful edge characteristics, therefore increasing the classification model's accuracy. Fig. 5 displays the particular structure of BAFEN.

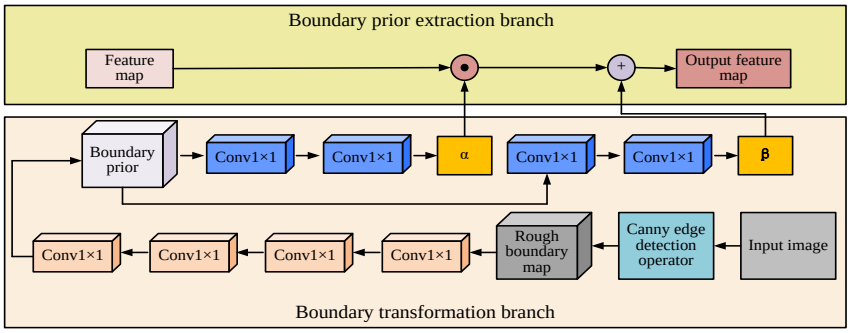


Fig. 5. BAFEN structure diagram.

In Fig. 5, the BAFEN structure mainly consists of two parts: the boundary prior extraction branch and the boundary transformation branch. Extracting features from boundary prior knowledge to serve as guidance for further processing is the primary task of the boundary prior extraction branch. The function of the boundary transformation branch is to transform the guidance information obtained from the boundary prior extraction branch into specific boundary features. First, Canny edge detection is performed on the input image, and then the detection result is used as a rough boundary map and a prior is applied to the boundary map. By extracting prior boundary knowledge and utilizing boundary transformation branches to generate the final boundary features. The mathematical calculation model of BAFEN is shown in Eq. (4).

$$BAFEN(F_i) = F_i \text{ e } a + b \tag{4}$$

In Eq. (4), F_i represents the output characteristics of the backbone network. a represents the boundary prior factor. b represents the boundary conversion factor. e represents Hadamard product. The range of values for F_i is shown in Eq. (5).

$$F = F_i \quad i \hat{I} \{1,2,3,4,5\} \tag{5}$$

In Eq. (5), F represents the set of output features. $i \hat{I} \{1,2,3,4,5\}$ represents the five CLs in BFEN. The calculation equation for a is shown in Eq. (6).

$$a = \text{Re}LU\left(Conv\left(\text{Re}LU\left(Conv\left(E_{prior}\right)\right)\right)\right) \tag{6}$$

In Eq. (6), $\text{Re}LU$ represents the activation function. E_{prior} represents the initial boundary map features. The calculation equation for b is shown in Eq. (7).

$$b = \text{Re}LU\left(Conv\left(\text{Re}LU\left(E_{prior}^{¢}\right)\right)\right) \tag{7}$$

In Eq. (7), $E_{prior}^{¢}$ represents the boundary map features after boundary transformation. By combining Eq. (6) and (7), the final boundary perception fusion feature can be obtained as shown in Eq. (8).

$$E_i = T_{conv}\left(Canny\left(F_i\right)\right) \tag{8}$$

In Eq. (8), T_{conv} represents continuous convolution operation. E_i represents boundary perception fusion features.

After obtaining fusion features and boundary perception fusion features, the study adopted a learning based approach to fuse the two features, thereby generating CBFFN. Firstly, pixel to pixel addition is used to preliminarily integrate information between C_i and E_i. Then, three CLs with 3 * 3 kernels and a step size of 1 are used for feature screening and extraction. Finally, S_{i+1} is overlaid on the output features of the last layer through residual learning to obtain a fused feature map. The calculation equation is shown in Eq. (9).

$$S_i = S_{i+1} \text{Å} T_{conv}\left(C_i \text{Å} E_i\right) \tag{9}$$

In Eq. (9), S_i represents the final fused feature map. To boost confidence, a loss function is added during the model training procedure. Eq. (10) displays the CFEN's content loss function.

$$L_C(P,T) = L_{SSIM}(P,T) + L_{IOU}(P,T) \tag{10}$$

In Eq. (10), L_C represents the content loss function of CFEN. L_{IOU} and L_{SSIM} represent the IOU loss function and SIIM loss function, respectively. P represents the classification result predicted by the model. T represents the true classification result of the model. Eq. (11) displays the BAFEN boundary loss function.

$$L_E(P,T) = L_{CBCE}(P,T_E) \tag{11}$$

In Eq. (11), L_E represents the boundary loss function of BAFEN. L_{CBCE} represents the CBCE loss function. T_E represents the boundary map of salient targets in real classification results. The fusion loss function of CBFFN is shown in Eq. (12).

$$L_F(P,T) = l_1\left(L_{IOU}(P,T) + L_{SSIM}(P,T) + l_2 L_{CBCE}(P,T)\right) \tag{12}$$

In Eq. (12), l_2 and l_1 both represent weight values. L_F represents the fusion loss function.

IV. Performance Testing and Application Analysis of Multi-level Feature Weighted Fusion Network Model for Domestic Garbage Image Classification

To verify the performance and application effectiveness of MFWFNM, this study compared three different image classification models: residual network (ResNet), faster region-based convolutional neural network (Faster R-CNN), and deformable convolutional networks (DCN). Using CA, classification error, and model iteration fitness values as performance evaluation indicators, it has been proven that the classification performance of MFWFNM is superior to other comparative models.

A. Performance Testing of Multilevel Feature Weighted Fusion Network Model

To demonstrate the benchmark performance of the model, this study chose a publicly available image dataset for simulation experiments. In order to avoid equipment errors during the experimental process, this study conducted comparative experiments in the same experimental environment. Table II displays the experimental setup and dataset details.

Table II provides the experimental environment and dataset information for this simulation experiment. Table II lists the three publicly accessible datasets that are chosen for this study's model performance testing: ImageNet, COCO, and TrashNet. Firstly, the iterative fitness changes of ResNet, Faster R-CNN, DCN, and MFWFNM are obtained as shown in Fig. 6.

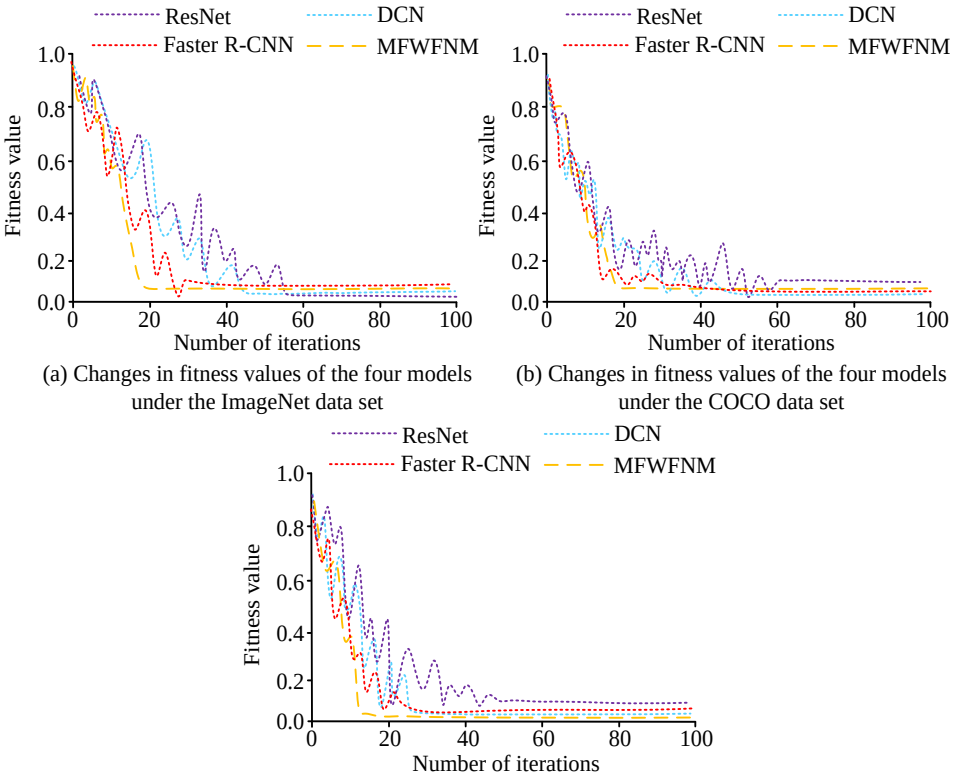
In Fig. 6, the iterative fitness values of four models, ResNet, DCN, Faster R-CNN, and MFWFNM, are shown in the ImageNet, COCO, and TrashNet datasets. As shown in Fig. 6(a), ResNet, DCN, Faster R-CNN, and MFWFNM can reach a stable state by iterating 56, 44, 28, and 17 times respectively in the dataset ImageNet. As shown in Fig. 6(b), ResNet, DCN, Faster R-CNN, and MFWFNM can reach a stable state after 61, 48, 41, and 18 iterations in the dataset COCO, respectively. In Fig. 6(c), ResNet, DCN, Faster R-CNN, and MFWFNM can reach a stable state in the dataset TrashNet after 49, 25, 27, and 13 iterations, respectively. In summary, MFWFNM can quickly iterate to a stable state in all three datasets, indicating that the processing efficiency of the model is higher.

TABLE II. Experimental Environment and Data Set Information Table

Project	Composition	Configuration
Lab environment	Processor	Intel Core i9
	Graphics processor	NVIDIA RTX 3080
	Memory	32GB RAM
	Storage	1TBSSD
	Operating system	Ubuntu 20.04
	Deep learning framework	TensorFlow
	Coding software	Python 3.8
Dataset information	ImageNet	1000+ image types
	COCO	More than 80 complex scene images
	TrashNet	6 image data sets for garbage classification

Fig. 7(a), 7(b), and 7(c) show the CA values of the four models in the ImageNet, COCO, and TrashNet datasets, respectively. Fig. 7 shows that while the CA of the MFWFNM model stays above 0.9, the CA values of ResNet, DCN, and Faster R-CNN all change to some amount as the number of samples grows. In Fig. 7(a), the highest CA of ResNet, DCN, Faster R-CNN, and MFWFNM models in the ImageNet dataset is 0.77, 0.84, 0.89, and 0.96, respectively. As shown in Fig. 7(b), the highest CA of ResNet, DCN, Faster R-CNN, and MFWFNM models in the COCO dataset is 0.78, 0.82, 0.91, and 0.95, respectively. As shown in Fig. 7(c), the highest CA of ResNet, DCN, Faster R-CNN, and MFWFNM models in the TrashNet dataset is 0.77, 0.79, 0.90, and 0.98, respectively.

Fig. 8(a) and 8(b) show the mean square error (MSE) and mean absolute error (MAE) of the four models in the TrashNet dataset, respectively. In Fig. 8(a), the highest MSE values for the ResNet, DCN, Faster R-CNN, and MFWFNM models are 4.28, 3.11, 1.95, and 0.72, respectively. In Fig. 8(b), the highest MAE values for the four models ResNet, DCN, Faster R-CNN, and MFWFNM are 4.73, 3.81, 2.69, and 1.15, respectively. Overall, the MFWFNM model has better error performance when processing multi sample image data, and therefore can more accurately complete image classification tasks.

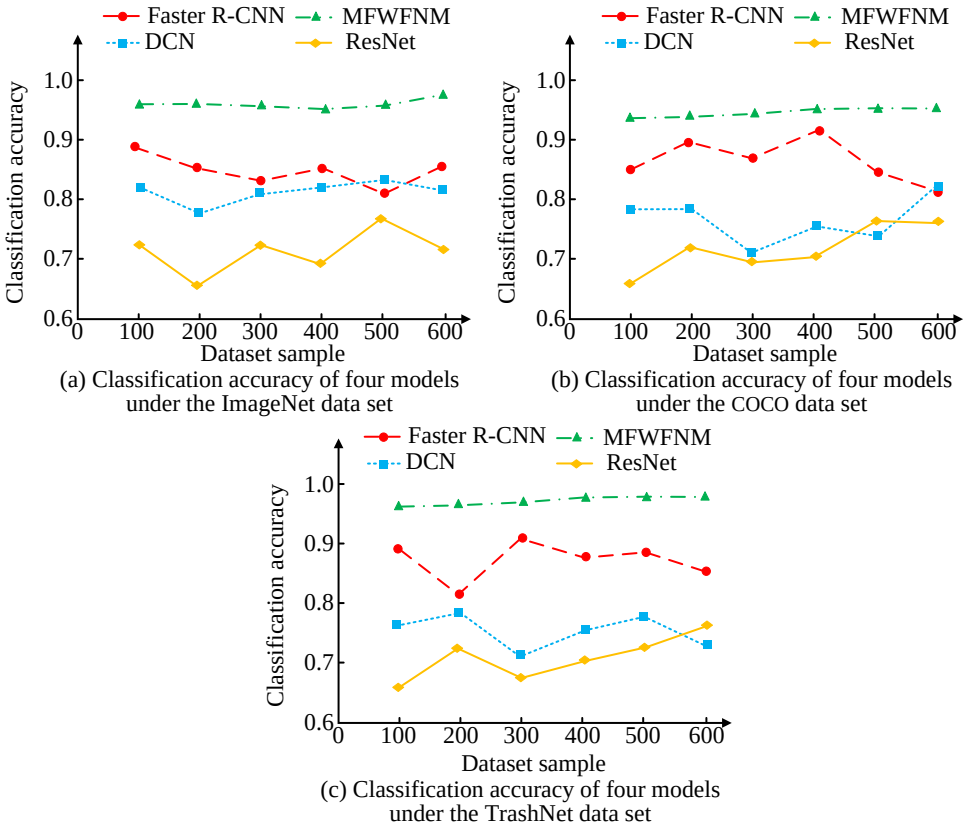


(a) Changes in fitness values of the four models under the ImageNet data set

(b) Changes in fitness values of the four models under the COCO data set

(c) Changes in fitness values of the four models under the TrashNet data set

Fig. 6. Iterative fitness values of four models in three data sets.

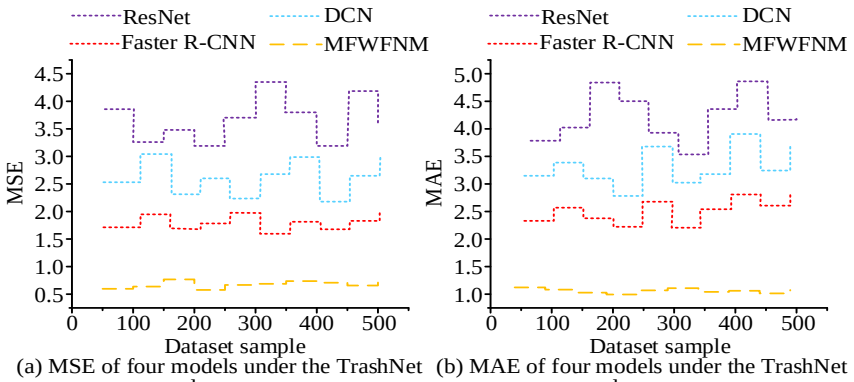


(a) Classification accuracy of four models under the ImageNet data set

(b) Classification accuracy of four models under the COCO data set

(c) Classification accuracy of four models under the TrashNet data set

Fig. 7. Classification accuracy values of four models in three data sets.



(a) MSE of four models under the TrashNet data set (b) MAE of four models under the TrashNet data set

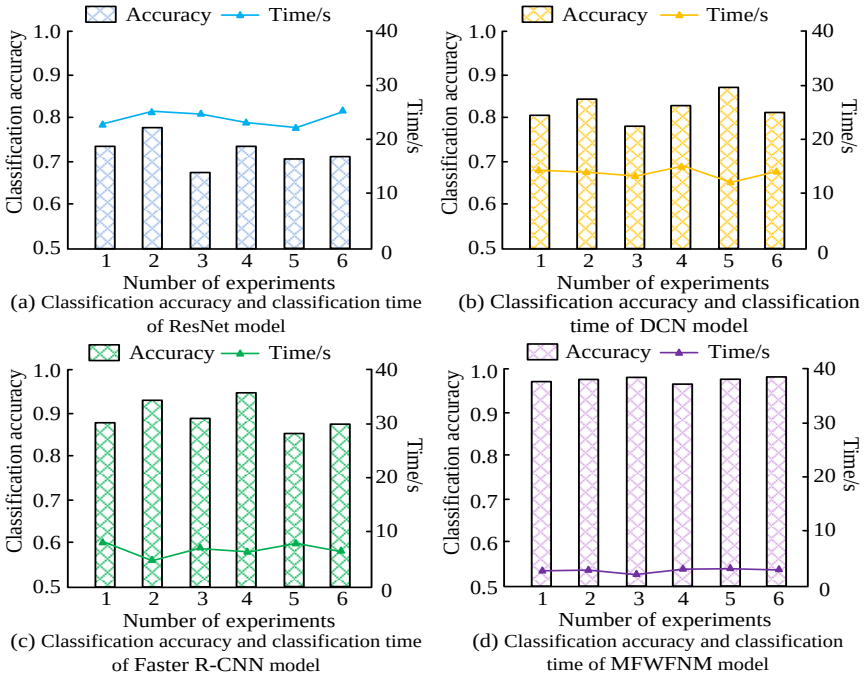
Fig. 8. Classification errors of the four models in the TrashNet data set.

B. *Application Effect of a Multilevel Feature Weighted Fusion Network Model in Actual Waste Classification*

Using the daily WC requirements listed in Table I, this study collects over 10,000 common DGIs to further validate the good performance of the MFWFNM model in real-world WC situations. These images are then used to test the actual classification performance of different classification models. 100 images of hazardous waste, recyclable waste, kitchen waste, and other waste are selected from over 10000 datasets, and the CA and time of the four models under multiple classification tests are shown in Fig. 9.

In Fig. 9, the CA and time consumption of four models, ResNet, DCN, Faster R-CNN, and MFWFNM, in actual garbage image classification are presented. In Fig. 9(a), in the six classification tests, the ResNet classification model achieved the highest CA of 0.77 and the shortest time consumption is 22 seconds. In Fig. 9(b), in the six classification tests, the DCN classification model achieves the highest CA of 0.86 and the shortest time consumption is 13 seconds. In Fig. 9(c), in the six classification tests, the Faster R-CNN classification model achieved the highest CA of 0.95 and the shortest time consumption is six seconds. In Fig. 9(d), the CA of the MFWFNM classification model can reach up to 0.98, and the shortest time required is only three seconds. Compared to the other three WC models, the MFWFNM model can achieve better classification performance in multiple classification tests.

In Fig. 10, the evaluation scores of experts and residents for the four classification models are presented. Assuming a total of five points, the higher the score, the higher the satisfaction. Based on Fig. 10(a), (b), (c), (d), experts and users have the highest satisfaction with the classification model MFWFNM, with scores concentrated in the first quadrant. On the contrary, the satisfaction of experts and users with the other three models is relatively scattered, but overall, the Faster R-CNN model has higher satisfaction scores than the other two models.



(a) Classification accuracy and classification time of ResNet model (b) Classification accuracy and classification time of DCN model

(c) Classification accuracy and classification time of Faster R-CNN model (d) Classification accuracy and classification time of MFWFNM model

Fig. 9. Actual classification accuracy and classification time of the four models in the garbage image classification problem.

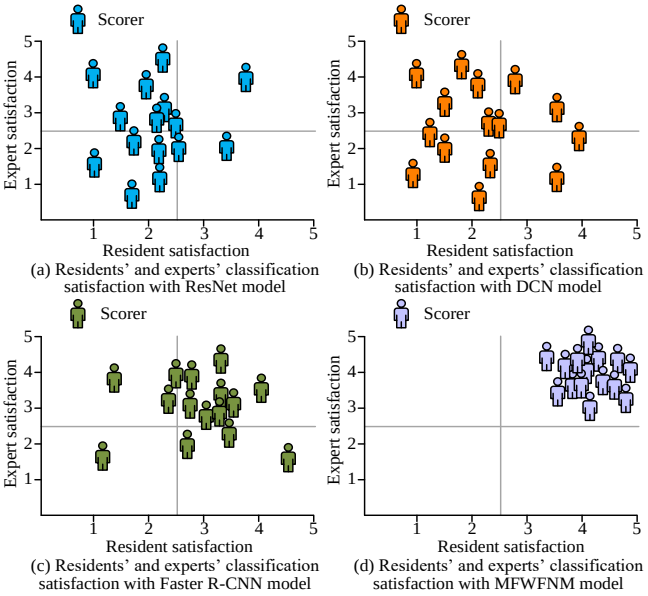


Fig. 10. Experts and residents' ratings of four classification models.

V. Discussion

In this study, the performance of MFWFNM model was analyzed, and the application effect of MFWFNM model in garbage classification was analyzed. The experimental results showed that MFWFNM model had significant advantages in CA, MSE and MAE. In the comparison experiment of MSE values, the highest MSE values of ResNet, DCN, Faster R-CNN and MFWFNM were 4.28, 3.11, 1.95 and 0.72, respectively. Moreover, the MFWFNM model proposed in this study had lower MSE values than the comparison model. The results showed that the introduction of VGG-16 network and the combination of four backbone networks improved the computational efficiency and accuracy of the model. Wang X et al. reached a similar conclusion in the relevant research of multi-feature fusion classification model [16]. Secondly, in the comparative analysis of CA and classification time, the highest CA and shortest classification time of MFWFNM model were 0.98 and 3s, respectively, which were lower than the comparison model. This result further indicated that, the fusion of BFEN, CFEN, BAFEN and CBFFN networks and the introduction of VGG-16 network improve the computational efficiency and image CA of MFWFNM model, thus improving the accuracy of garbage classification model. This conclusion coincided with the relevant conclusion proposed by Zheng Y et al in 2023 [17]. In terms of satisfaction evaluation, experts and users expressed the highest levels of satisfaction with the MFWFNM model proposed in the study, and its satisfaction ratings were significantly higher than those of the comparison model. This result was the same as the conclusion of the image classification model proposed by Xu X team [18]. The results showed that the proposed MFWFNM model had good practical value and was helpful to improve the efficiency and quality of garbage classification. In summary, compared with relevant studies, this study proposed the VGG-16 model as the main trunk network, combined with content and boundary perception model to build a multi-feature weighted fusion DGIC model. It has higher CA and precision in processing garbage images with complex backgrounds, diverse lighting conditions and changing shapes. This model not only provides convenience for accurately determining the types of domestic garbage, but also provides a certain idea and method for the theory of garbage image classification.

VI. Conclusion

Human daily life inevitably generates waste, and effective waste sorting and recycling are currently major research topics in the field. The present study utilized the VGG-16 network as the backbone and combined it with BFEN, CFEN, BAFEN, and CBFFN to create MFWFNM for DGIC. The findings revealed that MFWFNM outperforms ResNet, DCN, and Faster R-CNN in achieving a stable state, requiring only 17, 18, and 13 iterations in ImageNet, COCO, and TrashNet datasets, respectively, compared to the other models. Notably, MFWFNM's CA in all three datasets reached up to 0.96, 0.95, and 0.98, respectively, surpassing the accuracy of the compared model. Additionally, the MFWFNM model exhibited smaller classification errors with a highest MSE value of 0.72 and a highest MAE value of 1.15, as compared to ResNet, DCN, and Faster R-CNN models with the highest MSE values of 4.28, 3.11, and 1.95, and highest MAE values of 4.73, 3.81, and 2.69, respectively. Four classification models were utilized to categorize the DGI dataset outlined in the article. The MFWFNM classification model exhibited the highest CA of 0.98, with the shortest classification time of merely three seconds, and outperformed the other three models across all metrics. In summary, the study's WC model exhibits excellent classification performance. However, it is yet to be determined whether the model can generalize to other image types. Future research should investigate the classification performance of the model in diverse settings. This study has two shortcomings. First, the number of garbage types used in the classification test is limited. In contrast, the actual environment features an increasing number of complex garbage categories. Second, although a garbage classification model has been proposed, it lacks a systematic design and implementation. On the one hand, the exploration direction of future research is to collect more image data of household garbage to increase the data set of household garbage images, and to include existing garbage types as much as possible, so as to provide data basis for training classification models. Conversely, a convenient and optimal operating system is designed in accordance with the model and user requirements, thereby facilitating convenient and expedient services for pertinent users. Moreover, a multi-feature fusion algorithm and image classification method are employed to segment and recognize medical images, thereby enabling the effective extraction of the diseased area of patients and the recognition of images. It is therefore the objective of future research to enhance the efficiency with which doctors utilize medical images for disease analysis, thereby improving the efficiency of subsequent treatment.

References

[1] Sharma T, Nair R, Gomathi S. Breast cancer image classification using transfer learning and convolutional neural network. International Journal of Modern Research, 2022, 2(1): 8-16.

[2] Hosseinalipour A, Ghanbarzadeh R. A novel approach for spam detection using horse herd optimization algorithm. Neural Computing and Applications, 2022, 34(15): 13091-13105.

[3] Guo Y, Mustafaoglu Z, Koundal D. Spam detection using bidirectional transformers and machine learning classifier algorithms. Journal of Computational and Cognitive Engineering, 2023, 2(1): 5-9.

[4] Ahraminezhad A, Mojarad M, Arfaeinia H. An intelligent ensemble classification method for spam diagnosis in social networks. International Journal of Intelligent Systems and Applications, 2022, 14(1): 24-31.

[5] Yi N, Luo W. Research on Water Garbage Detection Algorithm Based on GFL Network. Frontiers in Computing and Intelligent Systems, 2023, 3(1): 154-157.

[6] Ma X, Li Z, Zhang L. An improved ResNet-50 for garbage image classification. Tehnički vjesnik, 2022, 29(5): 1552-1559.

[7] Wu Z, Li H, Wang X, Wu Z, Zou L, Xu L, Tan M. New benchmark for household garbage image recognition. Tsinghua Science and Technology, 2022, 27(5): 793-803.

[8] Vivekanandam B. Spam Email Classification by Hybrid Feature Selection with Advanced Machine learning Algorithm-Future Perspective. Journal of Soft Computing Paradigm, 2022, 4(2): 58-68.

[9] Li Y, Liu W. Deep learning-based garbage image recognition algorithm. Applied Nanoscience, 2023, 13(2): 1415-1424.

[10] Zou W, Zhang D, Lee D J. A new multi-feature fusion based convolutional neural network for facial expression recognition. Applied Intelligence, 2022, 52(3): 2918-2929.

[11] Yujiao Z, Ang L W, Shaomin S, Palaniappan S. Dropout Prediction Model for College Students in MOOCs Based on Weighted Multi-feature and SVM. Journal of Informatics and Web Engineering, 2023, 2(2): 29-42.

[12] Zhang G, Xu W, Gong H, Sun L, Li C, Chen H, Xiang D. Multi-feature fusion-based strabismus detection for children. IET Image Processing, 2023, 17(5): 1590-1602.

[13] Zhang J, Fang Z, Wang Z. Multi-feature fusion enhanced transformer with multi-layer fused decoding for image captioning. Applied Intelligence, 2023, 53(11): 13398-13414.

[14] Huang L, Li M, Xu T, Dong S. A waste classification method based on a capsule network. Environmental Science and Pollution Research, 2023, 30(36): 86454-86462.

[15] Amin S N, Shivakumara P, Jun T X, Chong K Y, Zan D L L, Rahavendra R. An Augmented Reality-Based Approach for Designing Interactive Food Menu of Restaurant Using Android, Artificial Intelligence and Applications. 2023, 1(1): 26-34.

[16] Wang X, Duan L, Ning C. Global context-based multilevel feature fusion networks for multilabel remote sensing image scene classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14(2): 11179-11196.

[17] Zheng Y, Li C, Zhou X, Chen, H., Xu, H., Li, Y. Application of transfer learning and ensemble learning in image-level classification for breast histopathology. Intelligent Medicine, 2023, 3(2): 115-128.

[18] Xu X, Huang L, Wu R, Zhang, W., Ding, G., Liu, L. Multi-feature fusion method for identifying carotid artery vulnerable plaque. IRBM, 2022, 43(4): 272-278.

Novel Biomarkers for Colorectal Cancer Prediction

Mohamed Ashraf[1], M.M. El-Gayar[2]*, Eman Eldaydamony[3]
Department of Information Technology, Faculty of Computers and Information Mansoura University, Mansoura 35516, Egypt[1, 2, 3]
Department of Computer Science, Arab East Colleges, Riyadh, Saudi Arabia[2]
Department of Information Technology, Faculty of Computers and Information Mansoura University, Mansoura 35516, Egypt[3]

***Abstract*—Most researchers work on solving the important issue of identifying biomarkers linked to a certain disease, like cancer, in order to assist in the disease's diagnosis and treatment. Several research have recently suggested several methods for identifying genes linked to disease. A handful of these methods were created specifically for CRC gene prediction, though. This research presents a novel prediction technique to determine new biomarkers related to CRC that can assist in the diagnosing process. First, we preprocessed four Microarray datasets (GSE4107, GSE8671, GSE9348 and GSE32323) using RMA (Robust Multi-Array Average) method to remove local artifacts and normalize the values. Second, we used the chi-squared test for feature selection to identify some significant features from datasets. Finally, the features were fed to XGBoost (eXtreme Gradient Boosting) to diagnose various test scenarios. The proposed model achieves a high mean accuracy rate and low standard deviation. When compared to other systems, the experiment findings show promise. The predicted biomarkers are validated through a review of the literature.**

***Keywords*—Colorectal cancer (CRC); microarray; biomarkers; gene expression omnibus; feature selection; chi-squared test; XGBoost**

I. Introduction

Particularly with the current state of treatment, cancer is a complex disease [1]. Colorectal cancer is one of the leading causes of cancer-related deaths worldwide. Colorectal cancer is a type of cancer that affects the latter part of the gut intestine within the digestive system [2]. Colorectal cancer is a cancer that occurs in the last 15 centimeters of the colon that meets part of the rectal region, and these two types of cancer together call colorectal cancer [3]. In most cases, colorectal cancer begins as a small mass of non-cancer cells called adenomatous polyp, after a period of time the solids that have formed become cancer masses present in the colon. These masses may be small and accompanied by very few symptoms [4]. There may be no symptoms of colorectal cancer, especially in its early stages. Common symptoms of colorectal cancer:

- Constipation or diarrhea.
- The feeling that the intestines are not completely emptied.
- Blood in the feces.
- Frequent pain caused by gases, bloating or feeling full.
- Weight loss for no reason.
- Persistent fatigue.
- Vomiting and nausea.

Colorectal Cancer is the third most common cancer in men and the second most common in women worldwide [2]. Despite recent advances in surgical and multimodal therapies, the overall survival of advanced CRC patients remains very low. Periodic tests can significantly reduce and prevent the incidence of this disease. If colorectal cancer is detected early enough, it can mostly be treated [5].

Determining the patient's cancer kind and its biomarkers [6] is one of the biggest problems facing researchers. The biomarker is the most crucial element in cancer research since it facilitates therapy and reduces the cost and time required for diagnosis. Thus, one of the researchers' most important tasks is to identify the most relevant biomarker. Because cancer involves dynamic genetic alterations, researchers have worked hard to investigate how to diagnose and forecast the disease.

Next-generation sequencing (NGS) and microarray data [7] are two important sources of potentially helpful molecular patterns. The abundance of gene expression data [8] facilitates the discovery of disease class and cancer-related biomarkers. Microarray technology can be used to study the entire genome, proteome, and transcriptome in various cells and tissues. The ability to analyze vast amounts of data quickly is one of the advantages of microarray technology.

In the context of microarray technology, feature selection techniques [9, 10] fall into three categories: filter, embedding, and wrapper techniques. The filter approach, which is independent of the predictor and does not require the classifier, assesses and ranks the genes in relation to the class label. Gene interactions and correlations are not taken into consideration. In contrast, the wrapper approach depends on whether features are added or removed while evaluating the subgroup features using classification methods. When compared to alternative classification algorithms, the filter approach is faster but less accurate than the wrapper method. In contrast, the wrapper approach is slower computationally but yields more accurate results than the filter method. The embedded approach involves building a particular classifier but uses search methods for subsets of optimal characteristics. The merging of the filter and wrapper techniques is intended to reduce the wrapper method's computational complexity issue.

This proposed model uses the chi-squared test [11] for feature selection to identify some significant features from the datasets for colorectal cancer classification. Then, the features were fed to XGBoost (eXtreme Gradient Boosting) [12] to diagnose different test cases and identify new biomarkers related to colorectal cancer that can help in the early diagnosis

of that disease. Finally, the results were validated using SHAP algorithm to explain the importance of the proposed biomarkers in identifying colorectal cancer and expose the significant biomarkers.

The main contribution of the proposed system is the identification of new biomarkers related to CRC that can assist in the early diagnosis of the disease.

For the reader's convenience, seven sections make up the remainder of this work. The relevant literature, existing flaws, and how the suggested approach gets around them are covered in Section II. In Section III, the materials and methods are explained. Section IV introduces the datasets, assessment measures, and results. The experimental results and novel biomarkers are discussed in Section V. By applying the SHAP algorithm to validate the results, Section VI illustrates the significance of novel biomarkers. Lastly, Section VII concludes and summarizes the plans for future work. Table I lists the used abbreviations in this paper for easy reference of the reader.

TABLE I. THE USED ABBREVIATIONS

CRC	Colorectal cancer	IG	Information Gain
RMA	Robust Multi-array Average	ANN	Artificial Neural Network
XGBoost	eXtreme Gradient Boosting	P-SVM	Penalized Support Vector Machine
SHAP	Shapley Additive Explanations	XAI	Explainable Artificial Intelligence
NGS	Next Generation Sequencing	DEGs	Differentially Expressed Genes
TCGA	The Cancer Genome Atlas	CAD	Colorectal Adenocarcinoma
GEO	Gene Expression Omnibus	GA	Genatic Algorithm
CatBoost	Categorical Boosting	MCC	Moffitt Cancer Center
PCA	Principal component analysis	X^2	Chi-squared
DT	Decision Tree	NB	Naive Bayes
TP	True positive	RF	Random Forest
FP	False positive	AB	Adaboost
LR	Logistic Regression	GBDT	gradient boosting decision tree
LGBM	Light gradient-boosting machine	GB	gradient boosting
SVM	Support Vector Machine	LDA	Linear Discriminant Analysis
FN	False negative	TN	True negative
SE	Sensitivity	SPC	specificity
TPR	True positive rate	FPR	False negative rate
LASSO	Least Absolute Shrinkage and Selection Operator	AUPR	Area under precision-recall
NCBI	National Center for Biotechnology Information	AUC	Area Under the Curve
GSNFS	Gene Sub-Network-based Feature Selection	ACC	Accuracy
mRMR	minimum Redundancy Maximum Relevance	SD	Standard Deviation

II. RELATED WORK

In the field of biology, predicting genes linked to a disease is regarded as an active research topic. Genes linked to these diseases have been found and predicted by many researchers; some of these studies have focused specifically on colorectal cancer. Table II shows a summary of the current studies. For example, Ahmadieh-Yazdi et al. [13] presented an approach to predict disease-related biomarkers using the TCGA dataset [14, 15] and GEO dataset [8]. First, they used LASSO and P-SVM methods as feature selection to identify the most relevant DEGs. DEGs frequently chosen by these techniques were chosen for additional analysis. Second, they applied the Multilayer Perceptron technique in conjunction with the Artificial Neural Network (ANN) method to evaluate the effectiveness of each method's gene selection in distinguishing primary samples from metastatic malignancies.

Maurya et al. [16] proposed a novel framework to identify genes associated with CRC using the TCGA dataset and GEO dataset. They used Boruta as a features selection method to select significant genes. The most relevant genes were then utilized to create an ML-based prognostic classification model with Random Forest classifier.

Li, S et al. [17] proposed a supervised learning framework based on deep learning (DeepCSD) to identify cancer subtypes. They designed a minimalist feed-forward neural network to capture the distinct molecular features in different cancer subtypes.

Al-Rajab et al. [18] proposed a framework that provides a two-step multi-filter hybrid model to select features for the classification of colon cancer was proposed. A mixture of the Information Gain (IG) and Genetic algorithms (GA) is used in the initial stage of feature selection. In order to lower the amount of genes and acquire more relevant genes, the second stage employs the minimum Redundancy Maximum Relevance (mRMR) filter approach. Utilizing appropriate machine learning techniques for further analyzing the data is the final step. With the suggested framework model, it was discovered that the SVM, Decision Tree, Naïve Bayes, and K-Nearest Neighbor classifiers provided accurate and promising results.

According to Shuwen et al. [19], the early detection of liver metastasis is crucial for improving the prognosis of patients with colorectal adenocarcinoma (CAD), and the utilization of a single biomarker in conjunction with a classification model has greatly enhanced the ability to predict the metastasis of various cancer types. There aren't many reports on CAD, though. Thus, the purpose of this study was to determine the most appropriate classification model for CAD patients with liver metastases and, using that model, investigate the gene's metastatic process. For the purpose of identifying the CAD system with liver metastases, the CatBoost model—which was constructed using 33 feature genes—exhibited the greatest classification performance.

TABLE II. A REVIEW OF A FEW RECENT STUDIES FOR COMPARISON

Study	Year	Analysis	Methodology	Dataset
Ahmadieh-Yazdi et al. [13]	2023	Identifying the disease-related biomarkers based on LASSO and P-SVM	LASSO, P-SVM, ANN	TCGA, GEO
Maurya et al. [16]	2023	Identifying genes linked to colorectal cancer using gene expression	Boruta, RF	TCGA, GEO
Li, S et al. [17]	2022	Predicting cancer subtypes based on DeepCSD	NN, Deep learning	GEO
Al-Rajab et al. [18]	2021	Identifying cancer cells based on two-stage approach	A hybrid of IG and GA, DT, KNN, SVM and NB	GEO
Shuwen et al. [19]	2020	Predicting biomarkers from classifier for liver metastasis of colorectal adenocarcinomas using machine learning models	LR, NN, SVM, RF, GBDT, Catboost.	GEO, MCC
Kozuevanich et al. [20]	2020	Predicting genes related to disease based on GSNFS	Correlation-based, Information gain, GSNFS	GEO
Yanke et al. [21]	2020	Identifying the disease-related biomarkers based on Random Forest and Deep learning	Random walk restart algorithm, RF, Deep learning	TCGA, GEO
Ram et al. [22]	2017	Identifying the disease-related biomarkers based on Random Forest	RF	GEO, Array Express databases

Kozuevanich et al. [20] proposed that the combination of Gene Sub-Network-based Feature Selection (GSNFS) and feature selection is very promising to identify biomarkers associated with CRC because it requires fewer subnetworks to build a classifier and provides a performance comparable to that of a full data set classifier.

Yanke et al. [21] proposed a model that used complex networks, machine learning methods and deep learning technology to look for probable genes linked to colorectal cancer in the following seven types of colorectal cancer data: LUAD, LUSC, UCEC, BRCA, COAD, HNSC, and KIRC. The signed random walk restart algorithm was employed in this suggested model to extract features. The random forest is the machine learning technique employed in this model as a colorectal cancer classifier. This model also makes use of deep learning technologies to look for putative colorectal genes and offer a novel method of diagnosing colorectal cancer.

Ram et al. [22] proposed a model that used the Random Forest algorithm to rank and select the genes needed to properly diagnose and treat cancer. While preserving its accuracy for prediction, the Random Forest method produced extremely tiny gene groups.

In conclusion, previous studies did not explore all the biomarkers associated with colorectal cancer, nor did they achieve the highest percentages in the disease classification process. To overcome the several limitations of the current studies, as mentioned above, we designed a novel prediction system that primarily identifies new biomarkers related to CRC based on microarray dataset that can assist in the diagnosing process and that is considered a very important advantage compared to previous studies. First, we preprocessed four Microarray datasets (GSE4107 [23], GSE8671 [24], GSE9348 [25] and GSE32323 [26]) using RMA (Robust Multi-Array Average) method27 to remove local artifacts and normalize the values. Second, we used the chi-squared test [11] for feature selection to identify some significant features from datasets. Finally, the features were fed to XGBoost (eXtreme Gradient Boosting) [12] to diagnose various test scenarios. The proposed model achieves a high mean accuracy rate and low standard deviation. When compared to other systems, the experiment findings show promise. The predicted biomarkers are validated through a review of the literature.

III. MATERIALS AND METHODS

The main contribution of the proposed system is the identification of new biomarkers related to CRC that can assist in the early diagnosis of the disease. Using the RMA (Robust Multi-Array Average) approach [27], we first preprocessed four Microarray datasets (GSE4107 [23], GSE8671 [24], GSE9348 [25], and GSE32323 [26]) to eliminate local artifacts and normalize the data. Secondly, we selected certain important features from datasets using the chi-squared test [11] for feature selection. XGBoost (eXtreme Gradient Boosting) [12] was then fed the features in order to diagnose several test scenarios. The model that has been suggested has a low standard deviation and a high mean accuracy rate. The experiment results show promise when compared to other systems. Based on a review of the literature, the expected biomarkers are confirmed.

The proposed prediction system is depicted in Fig. 1 with a unique four-step architecture. First, the preprocessing step comprises four steps: background correction, normalization, summarization and log2 transformation for removing local artifacts and noise. Second, the most relevant features are selected using the Chi-squared test [11] as a feature selection. Third, these proposed features are fed to the XGBoost algorithm [12] for identifying various test cases. In conclusion, we assess the proposed system's performance using five metrics, revealing encouraging outcomes in comparison to other methods. The following subsections contain more information on the suggested prediction system.

A. Preprocessing

In the preprocessing step, we prepared and improved the original dataset to feed it to the feature selection algorithm in order to improve our proposed system and obtain accurate results. The microarray datasets were preprocessed using RMA (Robust Multi-Array Average) approach [27] which involves three main steps: Background correction, Normalization, and Summarization. First, the microarray datasets were background corrected to remove local artifacts and noise so measurements aren't so affected by neighboring measurements. Second, we normalized these datasets to remove array effects so measurements from different arrays are comparable. Then, we

summarized the normalized datasets for combining probe intensities across arrays so the final measurement represents gene expression level. Finally, the log2 transformation was applied to the gene expression levels as almost all preprocessing methods return expression levels on log2 scale which is the approximately right scale.

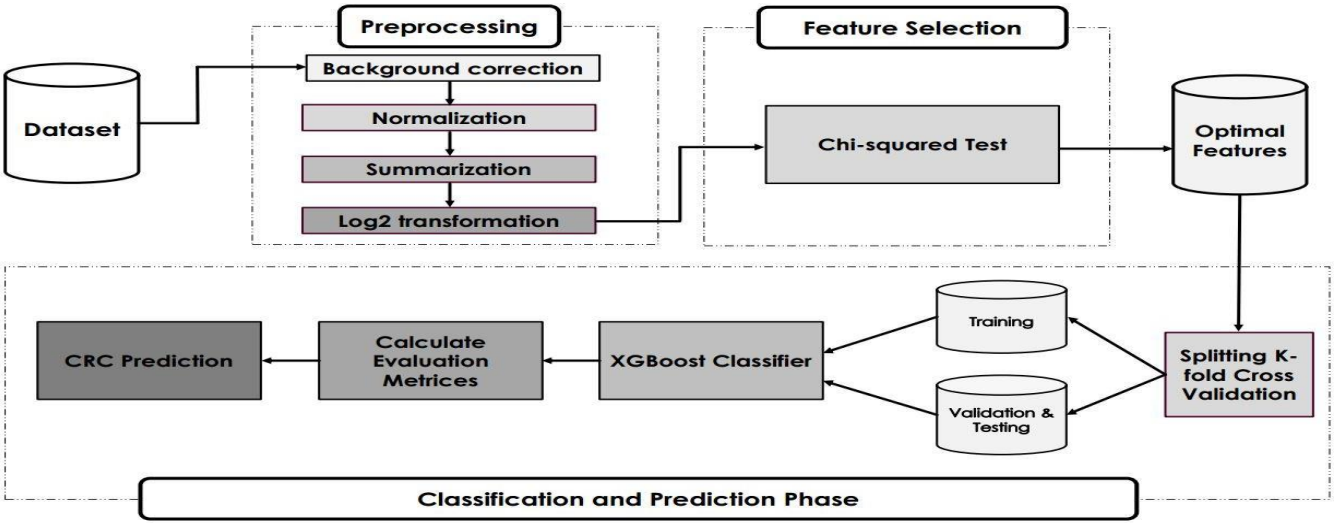


Fig. 1. The proposed prediction system.

B. Feature Selection

Feature selection stage seeks to reduce the set of features in microarray datasets by generating a new set of features from the preexisting ones. The majority of the data in the original set should be able to be summarized by these selected ones. This stage aids in lowering computation time, complexity, and model overfitting. Thus, we made an effort to narrow down the microarray datasets to the most important features. Suppose the classification algorithm receives incorrect or irrelevant features as input. In that scenario, it is unable to produce an accurate prediction because the machine learning model's performance depends heavily on the quality of the input data. As a result, we attempted to select the most notable features of most of the datasets. The selected features help us to correctly classify key genes associated with CRC. This is a crucial step in our proposed predictive model. If the features are not chosen correctly, the classification may be invalidated, impacting the predictive model.

There are three types of feature selection [9] in the microarray technology context: filter, wrapper and embedded. In the filter method, the genes are sorted and evaluated according to the class label. Correlation and gene-to-gene interactions are not considered, and it is independent of the predictor without using a learning algorithm (classifier). In contrast, the wrapper method depends on using the learning algorithm (classification algorithm) to add or remove features in order to evaluate the subset features.

This section outlines the feature selection method that produced encouraging results when compared to state-of-the-art methods: The Chi-squared test [11].

- Chi-squared test:

When large sample sizes are available, a statistical hypothesis test called a chi-squared test (also known as a chi-square [11] or X^2 test) is employed in the study of contingency tables. In layman's words, the main purpose of this test is to determine whether the two categorical variables (i.e., the two contingency table dimensions) have no effect on the test statistic. When the test statistic is chi-squared distributed under the null hypothesis, the test is considered valid. To ascertain whether there is a statistically significant discrepancy between the expected and observed frequencies in one or more categories of a contingency table, one can apply Pearson's chi-squared test. A Fisher's exact test is substituted for contingency tables with smaller sample sizes. The observations are categorized into classes that are mutually exclusive in the standard applications of this test. An X^2 frequency distribution is followed by the test statistic generated from the observations if the null hypothesis, which states that there are no differences between the classes in the population, is true. Assessing the observed frequencies' likelihood under the null hypothesis is the aim of the test. When the observations are independent, test statistics happen to follow an X^2 distribution. To verify if a pair of random variables are independent based on observations of each other, X^2 tests are also available. Datasets with categorical features are subjected to the Chi-square test. The desired number of features is determined with the optimal Chi-square scores by calculating the Chi-square between each feature and the target.

In 1900, Pearson published a paper on the X^2 test which is considered to be one of the foundations of modern statistics. In this paper, Pearson investigated a test of goodness of fit.

$$\sum_{i=1}^{k} p_i = 1 \quad (1)$$

$$\sum_{i=1}^{k} m_i = n \sum_{i=1}^{k} p_i = n \quad (2)$$

Suppose that n observations in a random sample from a population are classifed into k mutually exclusive classes with respective observed numbers of observations x_i (for i = 1,2,. . . ,k), and a null hypothesis gives the probability p_i that an observation falls into the *i*th class. So we have the expected numbers $m_i = np_i$ for all i.

$$x^2 = \sum_{i=1}^{k} \frac{(x_i - m_i)^2}{m_i} = \sum_{i=1}^{k} \frac{{x_i}^2}{m_i} - n \quad (3)$$

Pearson proposed that, under the circumstance of the null hypothesis being correct, as n → ∞ the limiting distribution of the quantity given above is the X^2 distribution as shown in Eq. (3). Pearson dealt first with the case in which the expected numbers *mi* are large enough known numbers in all cells assuming every observation *xi* may be taken as normally distributed, and reached the result that, in the limit as n becomes large, X^2 follows the X^2 distribution with *k* - 1 degrees of freedom.

To reduce the computational time and complexity for the classifier, the Chi-squared test is used to reduce the set of features and concurrently keep key features. We use Chi-squared test as feature selection to choose n features from the training model that have the highest score, based on threshold=5. After applying the Chi-squared test, we obtain 130 features instead of 54,675 features selected by Chi-squared test, as shown in Table III. As seen in Algorithm 1, we use the Chi-squared test to indicate the algorithm for the proposed feature selection after the preprocessing steps of RMA approach.

TABLE III. CHI-SQUARED FEATURE SELECTION AND THEIR NUMBERS

Dataset	# Of features before feature selection	# Of features after feature selection
GSE4107	54675	130
GSE8671	54675	130
GSE9348	54675	130
GSE32323	54675	130

Algorithm 1: The proposed preprocessing and feature Selection

Data: List of genes L_0

Result: The matrix of the most significant features F

Remove local artifacts and noise from the genes in L_0 and update it
Normalize the genes in L_0;
Represent probe intensities across arrays in L_0 as gene expression levels L_1
Transform the gene expression levels L_1 to log2 scale L_2
Initialize microarray matrix of features w
Apply the Chi-squared test for feature selection
Select the features with higher score than threshold=5 in matrix F

C. Classification

The selected features of the Chi-squared test are fed to the XGBoost classifer [12]. This classifier predicts the important genes linked to colorectal cancer (CRC) and diagnoses several test cases. Our tests demonstrate the superiority of XGBoost over state-of-the-art machine learning methods for both regression and classification tasks. XGBoost classifier is a machine learning algorithm that is applied for tabular and structured data. XGBoost is a fast and efficient implementation of gradient-boosted decision trees. XGBoost stands for extreme gradient boosting which implies that it is a large-scale machine learning technique with numerous components. XGBoost is an ensemble learning method. Relying just on the output of a single machine learning model may not always be adequate. Combining the prediction ability of several learners can be done methodically with ensemble learning [28-31]. All of the output from several models is aggregated into a single model as a result. The ensemble's models, often referred to as base learners, may come from different learning algorithms or from the same learning algorithms. The most popular ensemble learning models include bagging, boosting, stack generalization, and expert mixtures. However, two widely regarded methods for ensemble learning [28] are bagging and boosting. While these two methods can be used to a variety of statistical models, decision trees have been the most common application for them. The following is a summary of the main equations included in the XGBoost classifier:

1) Objective function: A regularized objective serves as the objective function in XGBoost and must be optimized throughout the training phase. It is composed of two terms: a regularization term that penalizes complexity to prevent overfitting, and a loss term that calculates the difference between the actual and predicted values.

$$obj(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} (f_k) \quad (4)$$

Where:

- $l(y_i, \hat{y}_i)$ is the loss function which measures the difference between the actual target value $\hat{y}_i$.
- (f_k) is the regularization term which penalizes the complexity of the model, where represents the k-th tree.
- θ represents the model parameters.

2) Gradient and hessian tree ensemble prediction: Gradient boosting is used in XGBoost to optimize the objective function. The first and second-order derivatives of the objective function with respect to the anticipated scores must be calculated in order to carry out gradient boosting.
Where:

$$g_i = \frac{\partial}{\partial \hat{y}_i} l(y_i, \hat{y}_i) \quad (5)$$

- g_i is the first-order derivative (gradient) of the loss function.

$$h_i = \frac{\partial^2}{\partial {\hat{y}_i}^2} l(y_i, \hat{y}_i) \quad (6)$$

- h_i is the second-order derivative (Hessian) of the loss function.

3) Tree ensemble prediction: The XGBoost model's final prediction is the weighted sum of the predictions made by several trees.

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \quad (7)$$

Where:

- $f_k(x_i)$ is the prediction from the k-th tree.

The XGBoost algorithm [12], which creates decision trees iteratively to minimize the above-defined objective function, is based on these equations. In numerous machine learning

applications, XGBoost provides state-of-the-art performance by optimizing the objective function. As seen in Algorithm 2, we illustrate the algorithm for the suggested classification based on the XGBoost classifier.

Algorithm 2: The proposed classification with XGBoost

Data: $D_{Train} = (x_1, y_1), (x_2, y_2), \dots, (x_N, x_N)$ $D_{Test} = (x_1, y_1), (x_2, y_2), \dots, (x_N, x_N)$
Result: Training model, and Prediction results
Use D_{Train} to train the XGBoost
Initialize the model as $h_0(x)$
For m = 1 → M do
  For i = 1 → N do
    Compute the loss function: $F(y, h_{m-1}(x_i))$
    Compute the residuals: rh,i
  End
  Fit a regression tree m_{th} to the $r_{h,i}$ values to create the terminal regions "tree leaf nodes" $R_{m,j}$, j = 1, 2, … , J; where J is the number of leaf nodes in the tree.
  For j = 1 → J do
    Get the v_m, j;
  End
  Update the weak classifier $h_m(x)$
End
Get the fnal model *H*(*x*)
Use D_{Test} to evaluate the prediction model

For s = 1 → N do
  Process XGBoost prediction model
  Get the predicted label
End
Utilizing the real and expected labels as inputs, compute the evaluation metric

IV. Experimental Results

This part includes the description of the datasets, the specifications for the hardware and software, and evaluation metrics and results. In the results subsection, first, we used the chi-squared test [11] for feature selection to identify some significant features from microarray datasets [7]. Second, the features were fed to XGBoost (eXtreme Gradient Boosting) [12] to diagnose various test scenarios. The proposed model achieves a high mean accuracy rate and low standard deviation. Third, when compared to other state-of-the-art classification algorithms, the experiment findings show promise. Finally, we use two performance measures [32] to provide some figures and tables that support a desired idea.

A. Datasets Description

Gene Expression Omnibus (GEO) [8] database served as the primary, accessible, and all-inclusive source for the gene expression raw data used in this paper when microarray data was deposited. The current study used four microarray data sets, as shown in Table IV, as follows:

- GSE4107 [23]: 22 samples and 54675 genes were included in this dataset. The data were split into two groups, each comprising 10 normal and 12 tumor samples.
- GSE8671 [24]: 64 samples and 54675 genes were included in this dataset. The data were split into two groups, each comprising 32 normal and 32 tumor samples.
- GSE9348 [25]: 82 samples and 54675 genes were included in this dataset. The data were split into two groups, each comprising 12 normal and 70 tumor samples.
- GSE32323 [26]: 34 samples and 54675 genes were included in this dataset. The data were split into two groups, each comprising 17 normal and 17 tumor samples.

TABLE IV. Description of the Datasets' Gene Expression

Dataset	Classification Type	# of Samples
GSE4107 [23]	normal	10
	tumor	12
GSE8671 [24]	normal	32
	tumor	32
GSE9348 [25]	normal	12
	tumor	70
GSE32323 [26]	normal	17
	tumor	17

B. Evaluation Metrics

This proposed work used five metrics [32] for measuring the performance of our proposed system, including accuracy (ACC), standard deviation (SD), precision, recall, and F1-Score, which are defined using Eq. (8) *to* Eq. (12).

$$ACC = \frac{TN+TP}{TN+FP+TP+FN} \quad (8)$$

$$SD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \quad (9)$$

- x_i being the result of the i-th measurement and $\bar{x}$ being the arithmetic mean of the n results considered.

$$Precision = \frac{TP}{FP+TP} \quad (10)$$

$$Recall = SEN = TPR = \frac{TP}{FN+TP} \quad (11)$$

$$F1 - score = \frac{TP}{TP+0.5(FN+FP)} \quad (12)$$

A true positive (TP) gene is one that has been accurately predicted to be a CRC gene, and this must be made clear. Reliability of genes correctly predicted as non-CRC genes is known as true negative (TN) rate. Erroneously predicted genes as colorectal cancer (CRC) genes are known as false positives (FP). False negative (FN) is also the rate of genes that are misclassified as not being CRC genes. The percentage of accurate results over all results based on TP and TN is known as the accuracy rate, or ACC. It evaluates how accurate the proposed system is. The precision can be defined as the ratio of correctly predicted results to the total number of wrong and accurate predictions, with "results" denoting the positive genes. The rate of correctly predicted results over all correctly

predicted results is known as the SEN, recall, or TPR, where "results" refers to the negative genes.

C. Results

All of the experimental results from this investigation, together with pertinent analysis, are provided in this subsection. The comparison of feature selection, the comparison of classification techniques, and the comparison with other prediction systems comprised the three parts of the experimental results.

1) Feature selection comparison: In the feature selection stage, we attempted to select the most notable features of most of the datasets. The selected features help us to correctly classify key genes associated with CRC. This is a crucial step in our proposed predictive model. If the features are not chosen correctly, the classification may be invalidated, impacting the predictive model. We employed the Chi-squared test to identify the important features from microarray datasets in order to construct our prediction system. The features from two state-of-the-art features (PCA and LASSO) are compared with the suggested features to validate them. When compared to other systems, the experiment findings show promise. We preformed the experiments based on microarray dataset using the XGBoost classifier with 5-fold and 10-fold cross-validation technique [33]. We evaluated the results using two performance metric [32]: **ACC** and **SD**.

Microarray dataset: Table V shows the performance comparison of the proposed features based on chi-squared test and features from state-of-the-art feature selection techniques: PCA and LASSO. For 5-fold cross-validation, the proposed features achieved the following: ACC equals 93.0% and SD equals 0.13. For 10-fold cross-validation, the proposed features achieved the following: ACC equals 91.8% and SD equals 0.14. The proposed features based on chi-squared test achieve a high mean accuracy and low standard deviation as shown in Table V.

TABLE V. THE PROPOSED FEATURES' EFFECTIVENESS WAS ASSESSED USING THE CHI-SQUARED TEST AND COMPARED TO TWO OTHER METHODS, NAMELY PCA AND LASSO WITH 5- AND 10-FOLD CROSS-VALIDATION USING THE MICROARRAY DATASET

Metric	K-fold	ACC (%)	SD
PCA	5 10	86.0 90.0	0.26 0.2
LASSO	5 10	82.0 80.0	0.15 0.22
Chi-squared test (proposed)	5 10	**93.0** **92.0**	**0.13** **0.14**

As shown in Fig. 2, the proposed features based on the chi-squared test achieve a high mean accuracy (ACC).

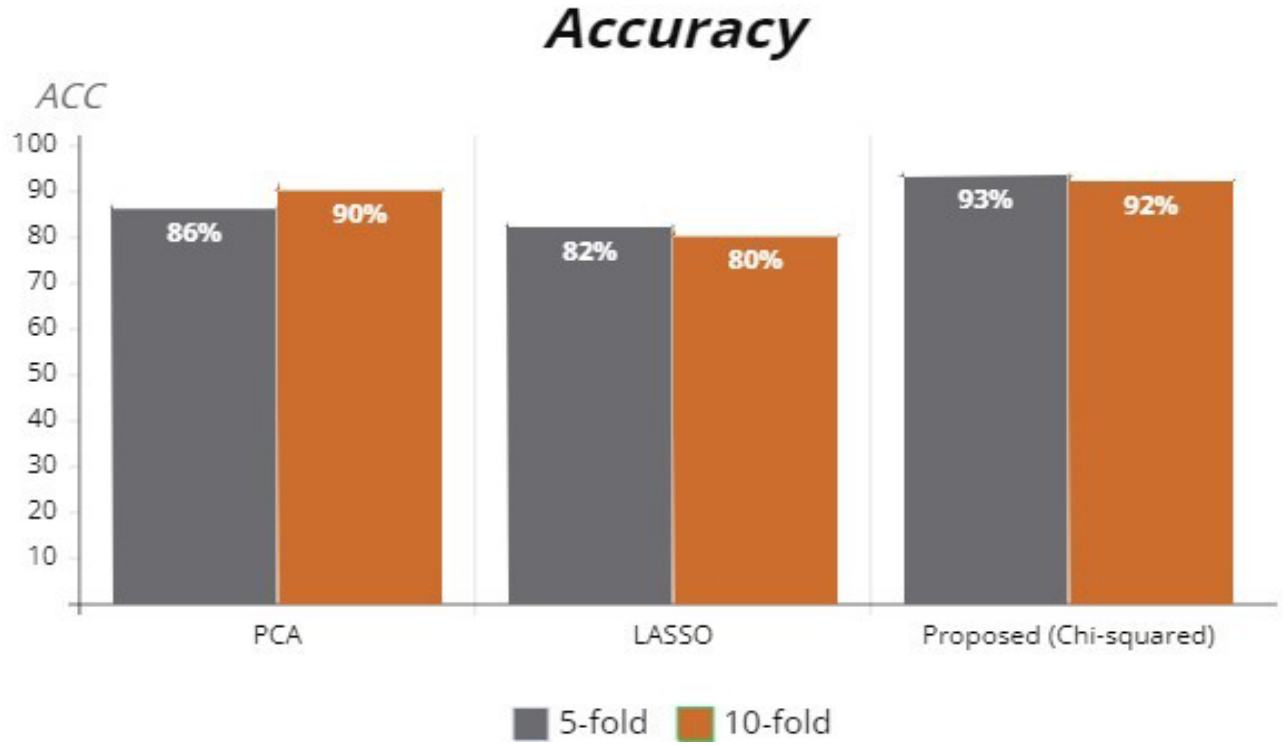


Fig. 2. The comparison between the proposed feature selection method and some state-of-the-art methods is based on accuracy.

2) Classification algorithm comparison: The selected features of the Chi-squared test are fed to the XGBoost classifier. This classifier predicts the important genes linked to colorectal cancer (CRC) and diagnoses several test cases. Our tests demonstrate the superiority of XGBoost over state-of-the-art machine learning methods for both regression and classification tasks. The results from three state-of-the-art classifiers (RF, LGBM, and LR) are compared with the XGBoost classifier to validate them. When compared to other systems, the experiment findings show promise. We

preformed the experiments based on four microarray datasets using the XGBoost classifier with 5-Fold and 10-fold cross-validation technique. We evaluated the results using five performance [32] metric: ACC, precision, recall, F1-Score and SD.

GSE4107 Microarray dataset: Table VI, shows the performance comparison of the proposed classifier based on XGBoost and classifiers from state-of-the-art techniques: LGBM, LR and RF. For 5-fold cross-validation, the proposed classifier achieved the following: ACC equals 93.0%, Precision equals 93.0%, Recall equals 100.0%, F1-score equals 96.0%, and Mean SD equals 0.085. For 10-fold cross-validation, the proposed classifier achieved the following: ACC equals 95.0%, Precision equals 90.3%, Recall equals 100.0%, F1-score equals 90.0%, and Mean SD equals 0.187. The proposed classifier based on XGBoost classifier achieved a high mean accuracy rate and low standard deviation as shown in Table VI.

As shown in Fig. 3, the proposed classifier based on XGBoost classifier achieved a high mean accuracy rate (ACC).

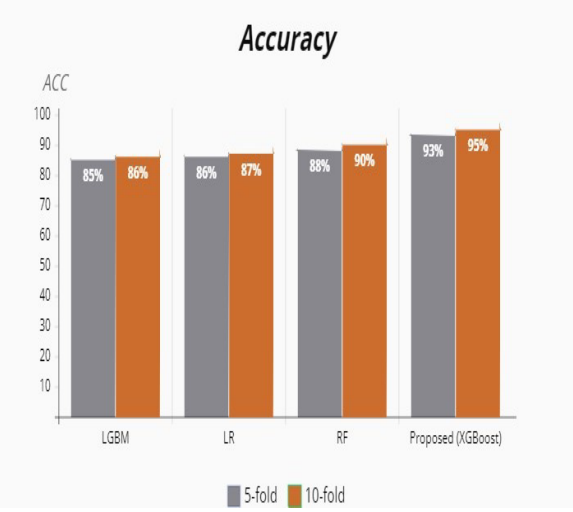


Fig. 3. The comparison between the proposed classifier and some state-of-the-art classifiers is based on accuracy using the GSE4107 dataset.

TABLE VI. USING 5-FOLD AND 10-FOLD CROSS-VALIDATION METHODS BASED ON THE 4107 DATASET, THE PROPOSED SYSTEM'S PERFORMANCE WAS EVALUATED IN COMPARISON TO STATE-OF-THE-ART CLASSIFIERS

Metric	K-fold	ACC (%)	Precision (%)	Recall (%)	F1-score (%)	SD
LGBM	5	85.4	87.2	89.5	82.4	0.180
	10	86.4	84.3	82.9	79.5	0.226
LR	5	86.2	89.6	86.3	87.4	0.155
	10	87.1	86.3	84.9	80.5	0.201
RF	5	88.9	90.2	94.5	91.4	0.120
	10	90.1	87.3	87.9	86.5	0.195
XGBoost	5	**93.0**	**93.0**	**100.0**	**96.0**	**0.085**
	10	**95.0**	**90.3**	**100.0**	**90.0**	**0.187**

GSE8671 Microarray dataset: Table VII, shows the performance comparison of the proposed classifier based on XGBoost and classifiers from state-of-the-art techniques: LGBM, LR and RF. For 5-fold cross-validation, the proposed classifier achieved the following: ACC equals 97.77%, Precision equals 96.0%, Recall equals 100.0%, F1-score equals 97.77%, and Mean SD equals 0.040. For 10-fold cross-validation, the proposed classifier achieved the following: ACC equals 98.0%, Precision equals 98.0%, Recall equals 100.0%, F1-score equals 98.88%, and Mean SD equals 0.037. The proposed classifier based on XGBoost classifier achieved a high mean accuracy rate and low standard deviation as shown in Table VII. As shown in Fig. 4, the proposed classifier based on XGBoost classifier achieved a high mean accuracy rate (ACC).

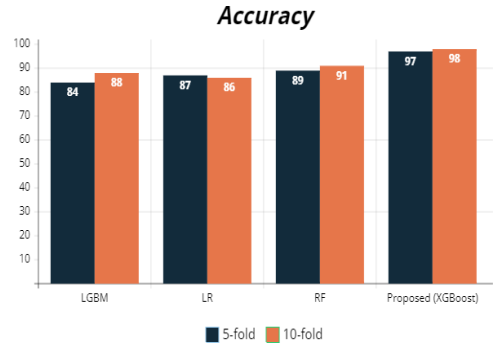


Fig. 4. The comparison between proposed classifier and some state-of-the-art classifiers based on Accuracy using GSE8671 dataset.

GSE9348 Microarray dataset: Table VIII, shows the performance comparison of the proposed classifier based on XGBoost and classifiers from state-of-the-art techniques: LGBM, LR and RF. For 5-fold cross-validation, the proposed classifier achieved the following: ACC equals 98.18%, Precision equals 100.0%, Recall equals 98.18%, F1-score equals 99.05%, and Mean SD equals 0.025. For 10-fold cross-validation, the proposed classifier achieved the following: ACC equals 91.0%, Precision equals 91.0%, Recall equals 100.0%, F1-score equals 95.05%, and Mean SD equals 0.057. The proposed classifier based on XGBoost classifier achieved a high mean accuracy rate and low standard deviation as shown in Table VIII. As shown in Fig. 5, the proposed classifier based on XGBoost classifier achieved a high mean accuracy rate (ACC).

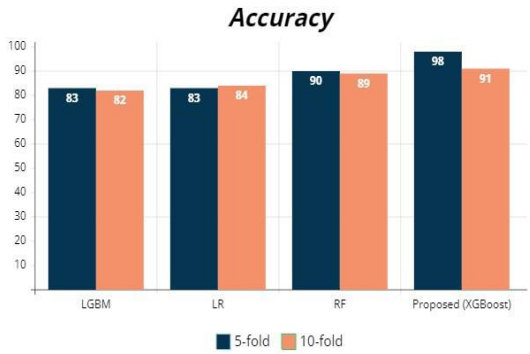


Fig. 5. The comparison between proposed classifier and some state-of-the-art classifiers based on accuracy using GSE9348 dataset.

TABLE VII. USING 5-FOLD AND 10-FOLD CROSS-VALIDATION METHODS BASED ON THE 8671 DATASET, THE PROPOSED SYSTEM'S PERFORMANCE WAS EVALUATED IN COMPARISON TO STATE-OF-THE-ART CLASSIFIERS.

Metric	K-fold	ACC (%)	Precision (%)	Recall (%)	F1-score (%)	SD
LGBM	5 10	84.2 88.5	85.4 85.6	85.5 81.4	84.1 76.4	0.190 0.236
LR	5 10	87.3 86.2	88.4 84.1	84.8 85.2	86.1 78.5	0.164 0.220
RF	5 10	89.4 91.4	91.1 89.5	93.4 88.4	90.2 88.2	0.130 0.188
XGBoost	5 10	**97.77** **98.0**	**96.0** **98.0**	**100.0** **100.0**	**97.77** **98.88**	**0.040** **0.037**

TABLE VIII. USING 5-FOLD AND 10-FOLD CROSS-VALIDATION METHODS BASED ON THE 9348 DATASET, THE PROPOSED SYSTEM'S PERFORMANCE WAS EVALUATED IN COMPARISON TO STATE-OF-THE-ART CLASSIFIERS.

Metric	K-fold	ACC (%)	Precision (%)	Recall (%)	F1-score (%)	SD
LGBM	5 10	83.8 82.1	84.1 81.2	84.2 80.2	81.2 76.8	0.240 0.311
LR	5 10	83.7 84.0	86.1 82.2	84.4 81.3	85.6 81.1	0.188 0.222
RF	5 10	90.1 89.5	92.3 89.3	95.4 88.9	92.7 87.8	0.116 0.187
XGBoost	5 10	**98.18** **91.0**	**100.0** **91.0**	**98.18** **100.0**	**99.05** **95.05**	**0.025** **0.057**

GSE32323 Microarray dataset: Table IX, shows the performance comparison of the proposed classifier based on XGBoost and classifiers from state-of-the-art techniques: LGBM, LR and RF. For 5-fold cross-validation, the proposed classifier achieved the following: ACC equals 96.0%, Precision equals 95.0%, Recall equals 100.0%, F1-score equals 97.14%, and Mean SD equals 0.059. For 10-fold cross-validation, the proposed classifier achieved the following: ACC equals 96.0%, Precision equals 100.0%, Recall equals 95.0%, F1-score equals 96.0%, and Mean SD equals 0.087. The proposed classifier based on XGBoost classifier achieved a high mean accuracy rate and low standard deviation as shown in Table IX.

TABLE IX. USING 5-FOLD AND 10-FOLD CROSS-VALIDATION METHODS BASED ON THE 9348 DATASET, THE PROPOSED SYSTEM'S PERFORMANCE WAS EVALUATED IN COMPARISON TO STATE-OF-THE-ART CLASSIFIERS.

Metric	K-fold	ACC (%)	Precision (%)	Recall (%)	F1-score (%)	SD
LGBM	5 10	89.6 88.8	89.4 85.4	88.7 86.6	86.6 81.4	0.160 0.216
LR	5 10	87.7 89.4	86.5 87.5	87.7 85.5	88.9 84.2	0.176 0.180
RF	5 10	92.8 91.5	92.5 91.5	95.5 93.9	96.5 94.5	0.105 0.125
XGBoost	5 10	**96.0** **96.0**	**95.0** **100.0**	**100.0** **95.0**	**97.14** **96.0**	**0.059** **0.087**

As shown in Fig. 6, the proposed classifer based on XGBoost classifer achieved a high mean accuracy rate (ACC).

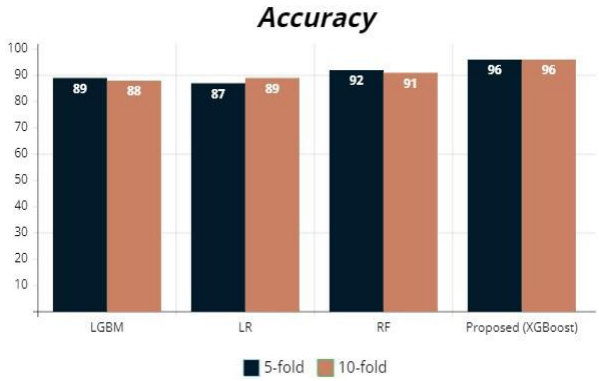


Fig. 6. The comparison between proposed classifier and some state-of-the-art classifiers based on Accuracy using GSE32323 dataset.

3) Comparison with other prediction systems: To verify how well the proposed system performs when using the XGBoost classification algorithm and the chi-squared feature selection technique. We evaluated how well the proposed system performed in comparison to state-of-the-art systems: Ahmadieh-Yazdi et al. [13], Maurya et al. [16], Li, S et al. [17], and Al-Rajab et al. [18].

The proposed system is compared to state-of-the-art systems in Table X. This comparison is based on feature selection, classification method, and performance evaluation using 10-fold cross-validation.

As shown in Fig. **7**, the proposed system achieved the highest mean accuracy rate (ACC), F1-score and Recall compared with state-of-the-art systems.

TABLE X. THE PROPOSED SYSTEM, BASED ON THE MICROARRAY DATASET, WAS COMPARED WITH THE PERFORMANCE METRICS, FEATURE SELECTION TECHNIQUES, AND CLASSIFICATION STRATEGIES EMPLOYED IN STATE-OF-THE-ART SYSTEMS

System	ACC (%)	F1-score (%)	Recall (%)	Classification Method	Feature Selection Method
Ahmadieh-Yazdi et al.[13]	90.0	80.9	84.6	ANN	LASSO
Maurya et al.[16]	99.0	80.2	83.6	RF	Boruta
Li, S et al.[17]	90.0	78.7	82.5	Deep learning	NN
Al-Rajab et al.[18]	DT=93.75,K-NN=93.75,NB=87.5,SVM=81.25	83.0	81.4	DT,K-NN,NB,SVM	a hybrid of IG and GA
Proposed System	**98.18**	**99.05**	**98.18**	**XGBoost**	**Chi-squared**

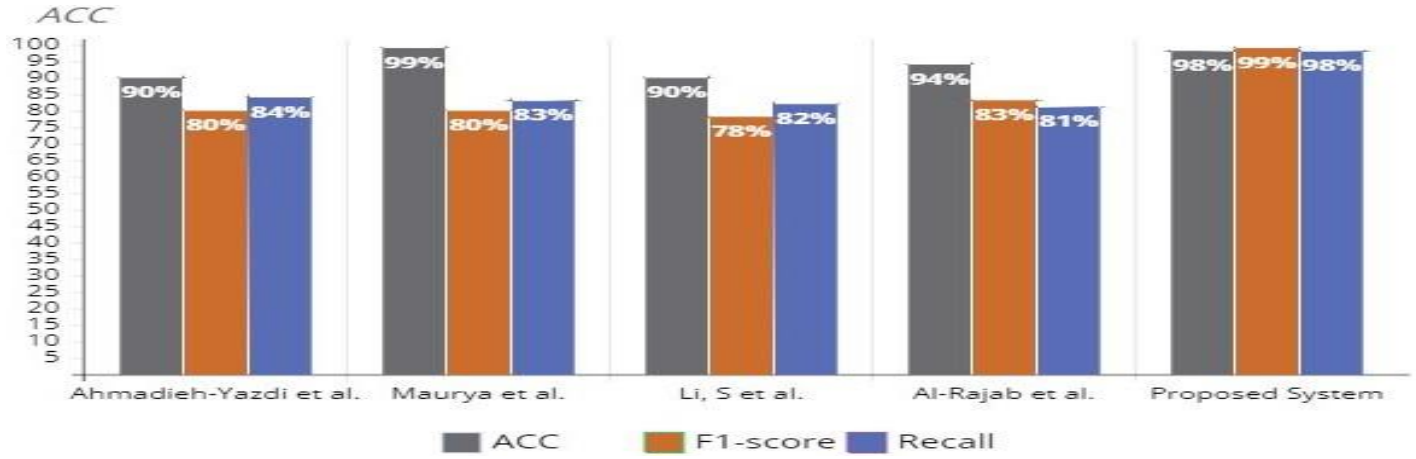


Fig. 7. The comparison between proposed system and some state-of-the-art systems.

V. DISCUSSION

Cancer is a complicated disease, especially in the treatment process so far. Colorectal cancer is one of the leading causes of cancer-related deaths worldwide. Colorectal Cancer is the third most common cancer in men and the second most common in women worldwide. Despite recent advances in surgical and multimodal therapies, the overall survival of advanced CRC patients remains very low. Periodic tests can significantly reduce and prevent the incidence of this disease. If colorectal cancer is detected early enough, it can mostly be treated. The biomarker is the most important component of cancer researches because it helps with the treatment process and saves cost and time in the diagnosis process. Therefore, determining the relevant biomarker is an essential task for the researchers. Researchers have made great efforts to explore an accurate diagnosis and prediction of cancer because it contains genetic changes that are dynamic. As a result, we developed our proposed prediction system to find key genes associated with colorectal cancer that can assist in the early diagnosis of the disease.

Using the RMA (Robust Multi-Array Average) approach, we first preprocessed four Microarray datasets (GSE4107, GSE8671, GSE9348, and GSE32323) to eliminate local artifacts and normalize the data. We employed 5-fold and 10-fold cross-validation methods to evaluate the proposed method. Secondly, we selected certain important features from datasets using the chi-squared test for feature selection. XGBoost (eXtreme Gradient Boosting) was then fed the features in order to diagnose several test scenarios. To verify how well the proposed system performs when using the XGBoost classification algorithm and the chi-squared feature selection technique. We evaluated how well the proposed system performed in comparison to state-of-the-art systems: Ahmadieh-Yazdi et al. [13], Maurya et al. [16], Li, S et al. [17], and Al-Rajab et al. [18]. The model that has been suggested has a low standard deviation and a high mean accuracy rate. The experiment results show promise when compared to other systems.

The results are evaluated using five performance metric: ACC, precision, recall, F1-Score and SD. For 5-fold cross-validation, the GSE4107 dataset achieved the following: ACC equals 93.0%, Precision equals 93.0%, Recall equals 100.0%, F1-score equals 96.0%, and Mean SD equals 0.085. For 10-fold cross-validation, the GSE8671 dataset achieved the following: ACC equals 98.0%, Precision equals 98.0%, Recall equals 100.0%, F1-score equals 98.88%, and Mean SD equals 0.037. For 5-fold cross-validation, the GSE9348 dataset achieved the following: ACC equals 98.18%, Precision equals 100.0%, Recall equals 98.18%, F1-score equals 99.05%, and Mean SD equals 0.025. For 5-fold cross-validation, the GSE32323 dataset achieved the following: ACC equals 96.0%, Precision equals 95.0%, Recall equals 100.0%, F1-score equals 97.14%, and Mean SD equals 0.059.

Lastly, the proposed model predicted novel CRC biomarkers that are not present in the databases [6] using the proposed prediction system. The literature review is used to verify these biomarkers. The Biomarkers that were extracted: VIP, CYR61, ADAMTS1, SLC51A, GREM1, PLN, MSI2, FOS, ADH1B, ETNK1, MEP1B, NR1H4, SYNP02, OGN, FOSB, UGT2A3, RGS1 and SERPINF1. We found out that

none of these genes had been linked to colorectal cancer (CRC) based on the literature review [6].

Table XI, show the smallest gene set for different datasets selected by XGBoost. These genes are known as colorectal cancer biomarkers and the function of each of these genes is presented in the mentioned table. The function and annotation of each selected gene were extracted from the NCBI database.

TABLE XI. THE LIST OF BIOMARKERS AND THEIR MAIN FUNCTIONS

Probe ID	Gene Symbol	Function
* "206577_*at*"	**VIP**	Encodes the vasoactive intestinal peptide (VIP) protein. VIP is a neuropeptide that has a wide range of physiological effects and is distributed widely throughout the body.
* "210764_*s_at*"	**CYR61**	Encodes a protein referred to as CCN1 (cysteine-rich protein 61),or cysteine-rich angiogenic inducer 61.
* "222162_*s_at*"	**ADAMTS1**	Encode the ADAMTS1 protein, an enzyme that is engaged in a number of biological functions, such as: Extracellular Matrix Remodeling, Angiogenesis Regulation, Cell Migration and Proliferation, and Tissue Homeostasis and Development.
* "228230_*at*"	**SLC51A**	Encodes the organic solute transporter alpha (OSTα) protein, which is a component of the heteromeric transporter complex that is involved in the transfer of bile acids.
* "218468_*s_at*"	**GREM1**	Encodes gremlin-1, a protein that belongs to the family of bone morphogenetic protein (BMP) antagonists known as DAN (differential screening-selected gene abnormal in neuroblastoma).
* "204939_*s_at*"	**PLN**	Phospholamban is essential for controlling the activity of the sarcoplasmic reticulum calcium ATPase (SERCA), a vital ion transporter.
* "1552364_*s_at*"	**MSI2**	Encodes the Musashi-2 (MSI2) protein, an RNA-binding protein belonging to the Musashi family.
* "209189_*at*"	**FOS**	Encodes the Fos protein, a member of the Fos transcription factor family.
* "209613_*s_at*"	**ADH1B**	Encodes the enzyme alcohol dehydrogenase 1B, which is essential to the liver's ethanol metabolism–the kind of alcohol present in alcoholic drinks.
* "224453_*s_at*"	**ETNK1**	Encodes the ethanolamine kinase 1 (ETNK1) enzyme, a component of the pathway responsible for phospholipid metabolism.
* "207251_*at*"	**MEP1B**	Encodes the beta subunit of meprins, zinc-dependent metalloendopeptidases composed of homo- and heterooligomers of 2 evolutionary related subunits, alpha (see MEP1A, 600388) and beta.
* "1554375_*a_at*"	**NR1H4**	Encodes a nuclear receptor called farnesoid X receptor (FXR).
* "225720_*at*"	**SYNP02**	Increases the resistance to immunotherapy and upregulates the infiltration of resting mast cells, which both contribute to the development of BLCA.
* "222722_*at*"	**OGN**	Encodes the osteoglycin protein, popularly referred to as mimecan.
"202768_*at*"	**FOSB**	Encodes the FosB protein, a member of the Fos transcription factor family.
"219948_*x_at*"	**UGT2A3**	Encodes UDP-glucuronosyltransferase 2A3, an enzyme belonging to the UGT family of enzymes.
"202988_*s_at*"	**RGS1**	Encodes the Regulator of G protein signaling 1 (RGS1) protein.
"202283_*at*"	**SERPINF1**	Encodes a protein called pigment epithelium-derived factor (PEDF), which belongs to the serpin superfamily of protease inhibitors.

VI. VALIDATION OF RESULTS

In this part, the importance of the proposed biomarkers are explained in identifying colorectal cancer by performing validation of results using SHAP algorithm [34, 35]. Model transparency, debugging, and fairness are all made possible by SHAP values, which offer a robust and theoretically valid way to understand each unique prediction of a machine learning model. Fig 8, shows the performance comparison of the proposed biomarkers based on XGBoost and other features using 4107 dataset.

Fig. 9 shows the performance comparison of the proposed biomarkers based on XGBoost and other features using 8671 dataset.

Fig. 10 shows the performance comparison of the proposed biomarkers based on XGBoost and other features using 9348 dataset.

Fig. 11 shows the performance comparison of the proposed biomarkers based on XGBoost and other features using 32323 dataset.

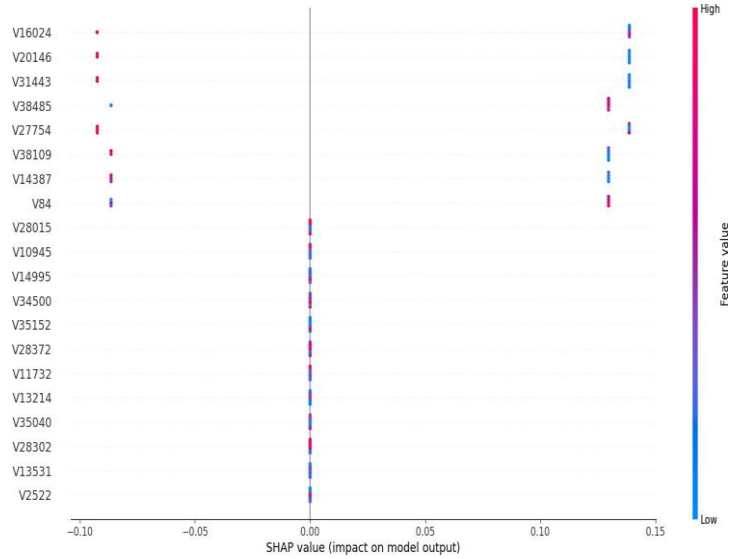


Fig. 8. The comparison between proposed biomarkers based on the 4107 dataset.

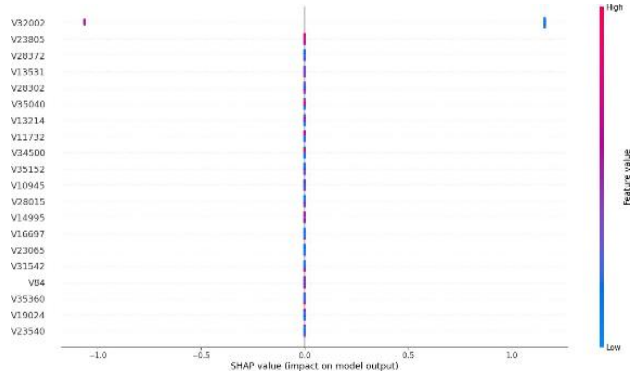


Fig. 9. The comparison between proposed biomarkers based on the 8671 dataset.

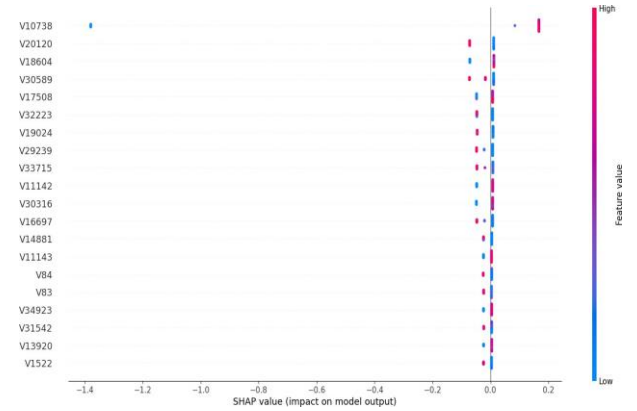


Fig. 10. The comparison between proposed biomarkers based on the 9348 dataset.

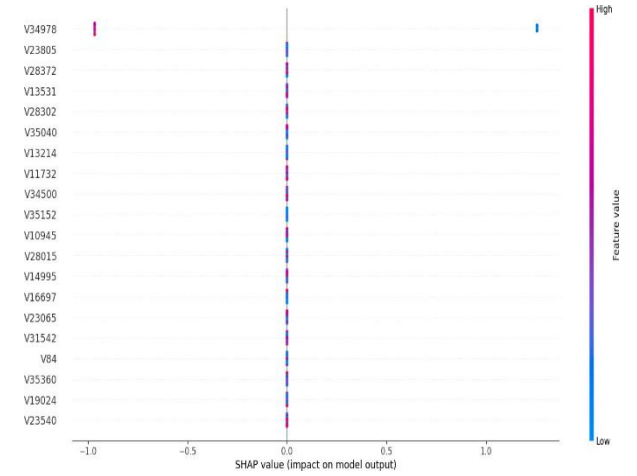


Fig. 11. The comparison between proposed biomarkers based on the 32323 dataset.

VII. Conclusion

This proposed work developed a novel prediction system to identify new biomarkers linked to CRC that can assist in an early diagnosis. The proposed model used four public microarray datasets: GSE4107, GSE8671, GSE9348 and GSE32323. The proposed prediction system comprises four steps. First, four Microarray datasets are preprocessed using the RMA (Robust Multi-Array Average) approach to eliminate local artifacts and normalize the data. Secondly, the certain important features from datasets are selected using the chi-squared test for feature selection. Then, the most relevant features were fed to XGBoost (eXtreme Gradient Boosting) to diagnose various test cases. Lastly, the results of the proposed system are assessed using five performance measures. The proposed model has a low standard deviation and a high mean accuracy rate. The experiment results show promise when compared to other systems. Based on a review of the literature, the expected biomarkers are confirmed. The future work is to find new biomarkers and gene alterations related to the different CRC grades. In the interim, the proposed system may be used to predict additional diseases that share similar genes.

VIII. Data Availability Statement

Yes, the model have research data to declare. Proposed model used four public Microarray datasets (GSE410723, GSE867124, GSE934825 and GSE3232326) downloaded from NCBI32 offcial Website.

References

[1] Yazdanpanah, N.; Rezaei, N. Interdisciplinary Approaches in Cancer Research. *Springer Nature* **2023**, pp. 1–16.

[2] American Cancer Society. *Colorectal Cancer Facts & Figures 2020-2022* (American Cancer Society, Atlanta, 2020).

[3] Smith, J. & Johnson, M. Colorectal cancer: A review. *Int. J. Cancer Res.* 10, 215–230, DOI: 10.1007/s00280-000-1234-5 (2000).

[4] Elshami, M. *et al.* Awareness of colorectal cancer signs and symptoms: a national cross-sectional study from palestine. *BMC Public Heal.* **22**, 866, DOI: 10.1186/s12889-022-13285-8 (2022). Accessed on April 30, 2022.

[5] Le, A., Salifu, M. & Mcfarlane, I. Artifcial intelligence in colorectal polyp detection and characterization. *Int. J. Clin. Res. & Trials* **6**, DOI: 10.15344/2456-8007/2021/157 (2021).

[6] Oh Hyung-Hoon, J. Y.-E. Novel biomarkers for the diagnosis and prognosis of colorectal cancer. *Intest Res* **18**, 168–183, DOI: 10.5217/ir.2019.00080 (2020). http://www.irjournal.org/journal/view.php?number=799.

[7] Hambali, M. A., Oladele, T. O. & Adewole, K. S. Microarray cancer feature selection: Review, challenges and research directions. *Int. J. Cogn. Comput. Eng.* **1**, 78–97, DOI: https://doi.org/10.1016/j.ijcce.2020.11.001 (2020).

[8] Clough, E. *et al.* NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. *Nucleic Acids Res.* **52**, D138–D144, DOI: 10.1093/nar/gkad965 (2023). https://academic.oup.com/nar/article-pdf/52/D1/D138/55039458/gkad965.pdf.

[9] Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A., Benítez, J. & Herrera, F. A review of microarray datasets and applied feature selection methods. *Inf. Sci.* **282**, 111–135, DOI: https://doi.org/10.1016/j.ins.2014.05.042 (2014).

[10] Veerabhadrappa & Rangarajan, L. Bi-level dimensionality reduction methods using feature selection and feature extraction. *Int. J. Comput. Appl.* **4**, 33 (2010).

[11] Pearson, K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos. Mag. Ser. 5* **50**, 157–175 (1900).

[12] Chen, T. & Guestrin, C. Xgboost: A scalable tree boosting system. *Proc. 22nd ACM SIGKDD Int. Conf. on Knowl. Discov. Data Min.* 785–794 (2016).

[13] Ahmadieh-Yazdi, A. *et al.* Using machine learning approach for screening metastatic biomarkers in colorectal cancer and predictive modeling with experimental validation. *Sci. Reports* **13**, 17, DOI: 10.1038/s41598-023-46633-8 (2023).

[14] Deepali, Goel, N. & Khandnor, P. Tcga: A multi-genomics material repository for cancer research. *Mater. Today: Proc.* **28**, 1492–1495, DOI: https://doi.org/10.1016/j.matpr.2020.04.827 (2020). International Conference on Aspects of Materials Science and Engineering.

[15] Liñares-Blanco, J., Pazos, A. & Fernandez-Lozano, C. Machine learning analysis of tcga cancer data. *PeerJ Comput. Sci.* **7**, e584, DOI: 10.7717/peerj-cs.584 (2021).

[16] Maurya, N. S., Kushwah, S., Kushwaha, S., Chawade, A. & Mani, A. Prognostic model development for classifcation of colorectal adenocarcinoma by using machine learning model based on feature selection technique boruta. *Sci. Reports* **13**, 14 (2023).

[17] Li, S. *et al.* Colorectal cancer subtype identifcation from differential gene expression levels using minimalist deep learning. *BioData Min.* **15**, 16 (2022).

[18] Al-Rajab, M., Lu, J. & Xu, Q. A framework model using multiflter feature selection to enhance colon cancer classifcation. *PLoS ONE* **19**, 26 (2021).

[19] Shuwen, H., Xi, Y., Qing, Z., Jing, Z. & Wei, W. Predicting biomarkers from classifer for liver metastasis of colorectal adenocarcinomas using machine learning models. *Cancer Medicine* **9**, 6667–6678, DOI: https://doi.org/10.1002/cam4.3289 (2020).

[20] Kozuevanich, S., Meechai, A. & Chan, J. H. Biomarker identifcation in colorectal cancer using subnetwork analysis with feature selection. *Springer Int. Publ.* **1149**, 119–127 (2022).

[21] Li, Y., Zhang, F. & Xing, C. Screening of pathogenic genes for colorectal cancer and deep learning in the diagnosis of colorectal cancer. *IEEE Access* **8**, 114916–114929, DOI: 10.1109/ACCESS.2020.3003999 (2020).

[22] Ram, M., Najaf, A. & Shakeri, M. T. Classifcation and biomarker genes selection for cancer gene expression data using random forest. *Iran. J. Pathol.* **12**, 339–347, DOI: 10.30699/ijp.2017.27990 (2017). https://ijp.iranpath.org/article_27990_2b14f68527d9aec085a93dbc82633079.pdf.

[23] Johnson, A. & Smith, B. GSE4107 dataset. Gene Expression Omnibus (GEO) (2020).

[24] Doe, John and Smith, Jane. GSE8671 dataset. Gene Expression Omnibus (GEO) (2008).

[25] Doe, John and Smith, Jane. GSE9348 dataset. Gene Expression Omnibus (GEO) (2010).

[26] Doe, John and Smith, Jane. GSE32323 dataset. Gene Expression Omnibus (GEO) (2012).

[27] Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M. & Hobbs, B. Summaries of affymetrix genechip probe level data. *Nucleic Acids Res.* **31**, e15, DOI: 10.1093/nar/gng015 (2003).

[28] Chan, H.-C., Chattopadhyay, A., Chuang, E. Y. & Lu, T.-P. Development of a gene-based prediction model for recurrence of colorectal cancer using an ensemble learning algorithm. *Front. Oncol.* **11**, DOI: 10.3389/fonc.2021.631056 (2021).

[29] F. E. Mohammed, N. S. Zghal, D. B. Aissa and M. M. El-Gayar, "Multiclassification Model of Histopathological Breast Cancer Based on Deep Neural Network," *2022 19th International Multi-Conference on Systems, Signals & Devices (SSD)*, Sétif, Algeria, 2022, pp. 1105-1111.

[30] Mohammed, F. E., Zghal, N. S., Aissa, D. B. & El-Gayar, M. M. (2022). Classify Breast Cancer Patients using Hybrid Data-Mining Techniques. *Journal of Computer Science*, *18*(4), 316-321.

[31] Lotfy, M.M., El-Bakry, H.M., Elgayar, M.M., El-Sappagh, S., I, G.A.M. et al. (2022). Semantic pneumonia segmentation and classification for covid-19 using deep learning network. *Computers, Materials & Continua*, *73(1)*, 1141-1158.

[32] Duda, R. O., Hart, P. E. & Stork, D. G. *Pattern Classifcation* (Wiley, 2012).

[33] Stephen Bates, T. H. & Tibshirani, R. Cross-validation: What does it estimate and how well does it do it? *J. Am. Stat. Assoc.* **0**, 1–12, DOI: 10.1080/01621459.2023.2197686 (2023). https://doi.org/10.1080/01621459.2023.2197686.

[34] Lundberg, S. M. & Lee, S.-I. A unifed approach to interpreting model predictions. (2017).

[35] National Center for Biotechnology Information. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/. Accessed: <date>.

Environmental and Economic Benefit Analysis of Urban Construction Projects Based on Data Envelopment Analysis and Simulated Annealing Algorithm

Jie Gong
School of Construction Management,
Chongqing Metropolitan College of Science and Technology, Chongqing, 402167, China

***Abstract*—With the continuous advancement of urbanization and the sustained growth of urban population, city building projects are facing severe challenges. How to analyze their environmental and economic benefits has become an urgent problem to be solved. Therefore, based on the proposed method for calculating the environmental and economic benefits of city building projects, this study uses a cross efficiency data envelopment analysis model for evaluation and solution. Then, an improved simulated annealing algorithm is used to achieve environmental and economic benefit optimization. The results showed that the improved simulated annealing algorithm tended to stabilize after 480 iterations, with maximum and minimum values of 0.86 and 0.21, respectively. The maximum F1 value was 0.988, indicating better performance. In the selected three urban construction projects, the cross efficiency data envelopment analysis model achieved high environmental and economic benefits, demonstrating the effectiveness of the model. After optimizing using the improved simulation degradation algorithm, the maximum economic benefit was increased by 850000 yuan, proving the effectiveness of the proposed method in analyzing the environmental and economic benefits of urban construction projects. It can provide more scientific decision support for construction project planning.**

Keywords*—DEA; simulated annealing algorithm; city building; environment; economics; benefit

I. INTRODUCTION

With the continuous improvement of urbanization level, China's construction industry has undergone technological innovation, and urban construction has developed rapidly [1]. During this process, many problems have arisen in the construction industry, especially sustainable development, which has had a huge impact on energy and the environment [2-3]. The reason is that urban buildings have a relatively long lifecycle, and their impact on the environment and energy consumption is also long-term, which makes it difficult for further expansion plans or existing renovations of the city to meet various requirements. Specifically, the construction unit needs to balance the impact on the surrounding environment while meeting the requirements for quantity and quality. Therefore, the analysis of environmental and economic benefits of urban construction projects has become a key research direction for many professionals. The existing analysis methods include cost-benefit analysis, life cycle assessment, regression analysis, etc. However, cost-benefit analysis is difficult to accurately quantify environmental impacts, and the evaluation of non-market factors is subjective. Although life cycle assessment methods can comprehensively consider environmental impacts, the implementation process is complex, data requirements are high, and it is difficult to provide decision support in a short period of time. Regression analysis relies on linear assumptions and is not suitable for handling nonlinear relationships. At the same time, it has high requirements for data and is easily affected by outliers. In recent years, Data Envelopment Analysis (DEA) models, Simulated Annealing (SA) algorithms, and multi-objective optimization provide theoretical and technical references for this direction [4-6]. In response to the prominent contradiction between the increase of water resources and the decrease of water supply, Wang Z et al. successfully constructed a multi-objective optimization configuration model using multi-objective programming theory. The results showed that the model had strong scientific practicality, which was of great significance for solving practical multi-objective optimization problems [7]. To solve the production demand fluctuations in unstable markets, Zhang Z et al. designed an optimized SERU model through a novel SERU production system that combined genetic algorithm and SA to achieve maximum profit increase. The results indicated that this model could effectively handle the SERU loading problem and had good robustness in solving the SERU loading model [8]. Gabi D et al. proposed a SA optimization scheme based on fruit flies to address the premature convergence of metaheuristic techniques and the imbalance between global and local searches. The scheme balanced local and global searches and was statistically analyzed using a 95% confidence interval. The results showed that the scheme could improve resource utilization by returning the minimum completion time and execution cost [9]. In response to the insufficient effectiveness and accuracy of grasshopper behavior research algorithms, Yu C et al. combined SA mechanism with original grasshopper optimization algorithm, and evaluated the relative ranking of this algorithm in CEC2017 through Friedman. The results showed that the proposed grasshopper optimization algorithm could effectively solve complex optimization problems and achieve results that meet or even exceed expectations [10]. In response to the shortcomings of previous design methods in

geothermal energy development, Liu J et al. proposed an innovative design framework that combined numerical simulation and SA, equipped with an energy lining system that generated thermal energy through circulating heat carrying fluids. The results indicated that the method still demonstrated strong processing ability under multiple factor crossover, which could guide practical applications [11]. In response to the slow convergence of existing slime mold algorithms in local search spaces, Ch L K et al. discovered the optimal solution of the objective function by mixing slime mold algorithms and SA algorithms to better change parameters. The non-convex, nonlinear, and typical engineering design difficulties were analyzed. The results showed that the mixed slime mold SA algorithm was more reasonable than other optimization techniques [12]. Camanho A S et al. conducted a literature review on the DEA of economic efficiency, including the optimization of costs, revenues, and profits. The application of different modeling methods was analyzed, providing new development directions for efficiency evaluation [13]. In response to the problem that traditional DEA cannot fully explore valuable information in big data, Zhu J focused on the development of nonlinear networks to propose corresponding DEA models. The results demonstrated the effectiveness of DEA in big data modeling [14].

By analyzing the research on DEA model and SA algorithm, both DEA model and SA algorithm have their own advantages, which are widely used in the fields of resource utilization efficiency and complex system optimization. Among them, the DEA model can effectively evaluate the relative efficiency of multiple decision-making units and is suitable for comprehensive evaluation of environmental and economic benefits. The SA algorithm can effectively search for global or approximate optimal solutions in the search space by simulating the physical annealing process. Its randomness and flexibility enable it to avoid getting stuck in local optima in complex multi-objective optimization problems. However, there is relatively little research on the combination of the two methods. Meanwhile, the economic input and output of urban construction projects have their own particularities, and the environmental impact also includes many complex factors. Therefore, starting from this point, the study combines DEA model and SA algorithm, proposes a method for calculating the environmental and economic benefits of urban construction projects, uses the cross efficiency DEA model for evaluation and solving, and combines the improved SA algorithm to optimize the environmental and economic benefits. Among them, the cross efficiency DEA model can better handle the relative efficiency evaluation between samples and reduce stability issues. The improved SA algorithm improves computational efficiency through optimization strategies and can quickly find approximate optimal solutions. The combination of these two methods can simultaneously consider economic and environmental benefits, providing a more comprehensive solution that is more suitable for the needs of complex urban construction projects.

The research content mainly includes four parts. The first part reviews the relevant research on DEA model and simulated annealing algorithm. The second part introduces the research methods. The first section analyzes the main factors in the environmental and economic benefits of urban construction projects and proposes calculation methods. The second section proposes a cross efficiency DEA model to solve multi-objective optimization problems of economic and environmental benefits. The third section proposes a combination of improved SA algorithm for optimizing economic and environmental benefits. The third part validated the performance of the improved SA algorithm and conducted empirical analysis on the environmental and economic benefits of urban construction projects. The fourth part summarizes and discusses the research results, and proposes future prospects.

II. Methods and Materials

In this chapter, the study first analyzes the main factors in the environmental and economic benefits of urban construction projects, and proposes calculation methods. Then, a combination of cross efficiency DEA model combined with improved SA algorithm is designed to optimize economic and environmental benefits.

A. Calculation of Environmental and Economic Benefits of Urban Construction Projects

The specific economic benefit indicators of urban construction projects are mainly divided into revenue indicators and cost indicators. Among them, the revenue indicators include the company's sales revenue, sales profit, and net profit. Cost indicators include material costs, labor costs, manufacturing costs, operating costs, and other aspects [15-16]. In the early stages of a construction project, costs mainly include market research, feasibility analysis, and technology development. Among them, urban buildings often face technical difficulties in the construction process. Technology research and development expenses are mainly investment related expenses for this research object. The construction cost can be divided into two parts: direct and indirect expenses. The labor, machinery, and material costs incurred during the construction project are considered direct expenses. The management fees, maintenance fees, and recycling fees of enterprises are classified as indirect expenses. Management fees refer to the property management fees incurred during the use of urban buildings, while maintenance fees are the maintenance costs incurred during the use process. The recycling cost includes the cost of dismantling components, cleaning the site, and processing fees for recycling, that is, the economic cost of dismantling, recycling, and utilization stages. In addition, the economic benefits of urban buildings include energy and resource conservation benefits, and energy conservation includes water and energy conservation. By calculating the unit electricity consumption throughout the life cycle of the building, the savings per square meter of urban buildings can be obtained, expressed as N_1. By consulting and referring to the local electricity standard P_1, the specific energy-saving benefits can be obtained, as shown in Eq. (1).

$$E_1 = N_1 * P_1 * A + S \tag{1}$$

In Eq. (1), A is the urban building area. S represents other energy saving costs. Combined with the local fixed price, the water-saving amount of urban buildings can be converted into water-saving benefits, as shown in Eq. (2).

$$E_2 = N_2 * P_2 \tag{2}$$

In Eq. (2), N_2 represents the water saving of urban construction projects. The benefits of resource conservation are mainly reflected in two aspects, namely the material saving benefits and time-saving benefits of urban construction. After comparing the differences in time efficiency between traditional and urban buildings for each floor, the time-saving benefits can be measured, as shown in Eq. (3).

$$E_3 = C * \left(\frac{D_c - D_z}{D_c} \right) \tag{3}$$

In Eq. (3), D_c and D_z represent the construction days of traditional and urban building projects, respectively. C represents the total cost of urban construction. Material saving benefits refer to the economic and environmental benefits obtained by optimizing resource utilization, improving material utilization efficiency, reducing raw material consumption and waste, and thereby reducing production costs and environmental burdens [17]. The specific calculation method is shown in Eq. (4).

$$E_4 = \sum N_i * P_i \tag{4}$$

A corresponding evaluation index system is created based on environmental benefits and costs to comprehensively assess the environmental impact and economic benefits of the model. Energy consumption includes water and electricity consumption, with water energy consumption referring to the amount of water resources used in the operation and model application of the power system. The energy consumption involves the amount of electricity consumed during model training and practical applications, including the computing resources required during model operation and the electricity consumed by equipment. Resource consumption includes the amount of wood, steel, and other resources consumed. Wood consumption involves the used wood resources in model development and system construction processes. Steel consumption includes the steel resources used in equipment manufacturing, infrastructure construction, and other aspects. The specific indicator system is shown in Fig. 1.

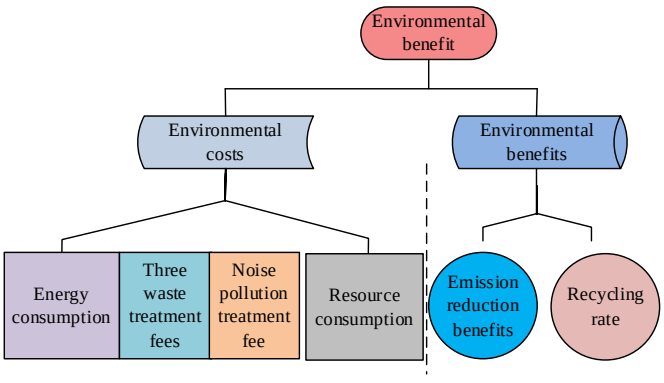


Fig. 1. Evaluation index system based on environmental benefits and environmental costs.

In order to fully evaluate the environmental impact of the project, the cost of waste treatment cannot be ignored. The cost of three wastes treatment refers to the cost of treating the wastewater, exhaust gas, noise, and solid waste generated during urban construction projects. The wastewater treatment fee covers the cost of wastewater treatment and discharge generated during construction and production processes, including the cost of wastewater collection, treatment, purification, and standard discharge. The cost of exhaust gas treatment mainly targets the exhaust gas generated during the construction process and vehicle exhaust. Usually, the exhaust emissions during the construction process do not exceed national standards, so there is less demand for treatment. The cost of solid waste disposal involves the disposal of solid waste generated during construction and production processes, such as construction waste, material residues, etc. The cost of noise pollution control includes the expenses required for implementing noise reduction measures, such as noise isolation and equipment maintenance. In the noise detection of urban buildings, the evaluation is mainly based on the relevant regulations of the urban area, as shown in Table I. The noise control can be assessed by measuring the residents' perception of noise in some areas.

TABLE I. URBAN REGIONAL NOISE ASSESSMENT STANDARDS

Adjacent	Distance/m
A type of standard applicable area	40~50
Applicable areas of Class II standards	23~35
Applicable areas of Class II standards	15~25

To reduce particulate matter emissions, the study adopts a five point sampling method, using a TH-150AH sampler to collect air samples and measure the particulate matter content in these samples. Then, the measured particulate matter content is compared with the relevant comprehensive standards. According to estimates, each square meter of dust reduction can bring 6.1 yuan in ecological benefits. Measures to reduce particulate matter not only help reduce environmental pollution, but also generate certain economic value. The specific income expression is shown in Eq. (5).

$$F = f * A * B \tag{5}$$

In equation (5), f represents the reduction of air particulate matter. B represents the ecological benefits of dust reduction in prefabricated buildings.

B. Cross Efficiency DEA Model

The study analyzed the main factors in the environmental and economic benefits of urban construction projects and proposed corresponding calculation methods. Next, it is necessary to solve the multi-objective optimization problem of economic and environmental benefits. As a non-parametric evaluation method, DEA is easy to calculate and can effectively process multi output and multi input data [18]. By creating an optimized model, it is possible to define an effective Pareto front and evaluate the efficiency of its object. In addition, the best weight combination among each decision unit can also be obtained. Therefore, when dealing with the subjective problems

caused by decision-makers' likes and dislikes in multi-objective programming, combining with multi-objective optimization solutions can be relatively easy to solve. Although traditional DEA models can effectively obtain decision units, they can only work on one decision unit, and the ranking of each decision unit depends on self-evaluation, resulting in significant errors. In response to the above two points, a method of cross evaluating its efficiency value is applied. According to the analysis, although the traditional DEA model inevitably lacks extremism, it is more suitable for dealing with practical situations [19]. In addition, traditional DEA models output decision units with many efficiency values of 1, which makes it difficult to distinguish whether they are of high quality. Cross efficiency DEA can further and hierarchically divide the decision units constructed by all efficiency values, and then obtain the rankings of all decision units. The cross efficiency DEA used in the study first calculates the negative ideal solution corresponding to the target value, including two maximization and minimization objectives. The negative ideal solutions for maximizing and minimizing the objectives are shown in Eq. (6).

$$\begin{cases} a_{ij}^{\min} = \min\left\{a_{i1}, a_{i2}, ..., a_{ij}\right\}, i = 1, 2 \\ b_{ij}^{\max} = \max\left\{a_{i1}, a_{i2}, ..., a_{ij}\right\}, i = 3, 4 \end{cases} \tag{6}$$

In Eq. (6), a_{ij} represents the maximization objective. b_{ij} represents minimization objective. Subsequently, the target values of each scheme are subtracted from the corresponding negative ideal solutions in all schemes to obtain the indicator values. The indicator value of the j-th scheme is Y_j. The conversions of a_{ij} and b_{ij} are shown in Eq. (7).

$$\begin{cases} y_{ij} = a_{ij} - a_{ij}^{\min}, i = 1, ...o \\ y_{ij} = b_{ij}^{\max} - b_{ij}, i = 3, 4, \ j = 1....o \end{cases} \tag{7}$$

After conversion, a set of homogeneous indicators can be obtained, which are then passed into the cross efficiency DEA to identify objective and effective solutions. After evaluating and analyzing the cross efficiency DEA, the efficiency ranking is then carried out. The top ranked solution is the optimal solution. For each evaluated multi-objective optimization scheme, since the objective transformation no longer involves input-output relationships, the input is usually considered as 1. Under the condition of unchanged scale benefits, applying the Charnes-Cooper transformation can obtain a linear programming model, as shown in Eq. (8).

$$\max E_{dd} = \sum_{i=1}^{4} \mu_{id} y_{id} - \delta \tag{8}$$

In Eq. (8), d is the evaluation scheme. E_{dd} is its self-evaluation relative efficiency. By solving the model, the optimal solution for scheme d is obtained. Through the evaluation and analysis of this optimal solution, the cross efficiency value of the scheme is obtained, as shown in Eq. (9).

$$E_{dj}^{*} = \frac{\sum_{i=1}^{4} \mu_{id}^{*} y_{ij}}{1 - \delta^2} (d, j = 1, ..., r) \tag{9}$$

After calculating the cross efficiency matrix, the arithmetic mean method is used to obtain the cross efficiency value, as shown in Eq. (10).

$$E_j^{*} = \frac{1}{n} \sum_{d=1}^{n} E_{dj}^{*} (n = r, j = 1, ..., r) \tag{10}$$

Among them, the higher the efficiency value, the more reasonable and excellent the solution is. By sorting efficiency values according to demand, the most effective solution can be identified, presenting decision-makers with an optimal solution that is not influenced by subjective factors. The application process of the cross efficiency DEA method is shown in Fig. 2.

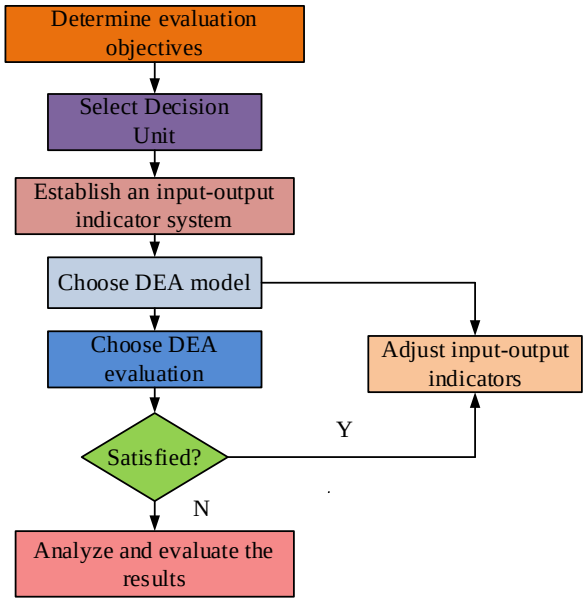


Fig. 2. Application process of cross efficiency DEA method.

C. Benefit Optimization Combined with Improved SA Algorithm

After using the cross efficiency DEA model for evaluation and solution, the study further adopts an improved simulated annealing algorithm to optimize environmental and economic benefits. SA algorithm is based on the basic principle of solid annealing and belongs to a probability-based algorithm [20-21]. Specifically, in the solid annealing, the solid is first heated to a certain degree and then slowly cooled. Due to heating, the particles inside the solid appear disordered, and the corresponding internal energy gradually increases. However, slow cooling gradually makes the particles more ordered, and equilibrium states are achieved at all temperature levels. Finally, when reaching room temperature, it becomes the ground state, and the internal energy decreases to the minimum [22]. The heating process, isothermal process, and cooling process together constitute the solid annealing process, which can actually simulate combinatorial optimization problems. That is, the solution space of the problem corresponds to the internal state of the object, the optimal solution corresponds to the lowest energy state, the objective function corresponds to the energy, and the initial temperature setting corresponds to the melting process. The Metropolis criterion optimizes the isothermal process, and the parameter controls the temperature

drop corresponding to the cooling process. The corresponding relationship is shown in Fig. 3.

The SA algorithm can escape from local optima until the global optimum is found. The calculation process is simple, making it highly applicable to combinatorial optimization problems. However, this algorithm is limited by the parameter of cooling rate, often resulting in significant time waste or the optimal solution being skipped, and has a long convergence time. Therefore, the study first introduces genetic algorithm to avoid getting stuck in local optima. A mixed mutation operator is proposed to divide the mutation process into insertion mutation, reversal mutation, and swapping mutation, reducing the randomness in the mutation process. Secondly, the Metropolis criterion is inserted to determine the entry of paternal chromosomes into the offspring population. The higher the fitness value, the greater the probability of passing. Based on these two operations, an improved SA algorithm is proposed, which can ensure the overall optimal solution results and accelerate the solving process. The improved SA algorithm first performs chromosome encoding using real number encoding. Then, real numbers in s rows and t columns that match the actual meaning are randomly obtained. Among them, t is the number of genes on the chromosome, and s is the initial population size. This generates the initial population. Afterwards, the selection operation is performed. To ensure the completeness of the optimal individual, the optimal preservation strategy, namely the elite preservation strategy, is adopted, combined with the roulette wheel method for joint screening. The elite retention strategy can enable the best individual to enter their offspring and conduct selective selection among secondary individuals, which can further place them on the offspring chromosome. Based on the roulette wheel selection method, individuals with lower fitness have a significantly increased probability of entering offspring. The individual's selection probability is shown in Eq. (11).

$$P_{selected} = \frac{fitness(a_i)}{\sum_{i=1}^{n} fitness(a_i)} \tag{11}$$

In Eq. (11), a_i represents an individual. $P_{selected}$ is the probability of being selected. $\sum_{i=1}^{n} fitness(a_i)$ is the sum of fitness values. After screening excellent paternal chromosomes, crossover operations can be carried out. In this process, it is necessary to exchange the two selected paternal chromosomes, so that offspring can inherit the excellent characteristics of the paternal chromosomes. To ensure the efficiency of the calculation, the single point crossing method is used in the study. A point on the crossing point of the parental chromosome is randomly selected, and two parental chromosomes are subjected to cleavage treatment to obtain the offspring chromosome after replacement. The effect before and after single point crossing is shown in Fig. 4.

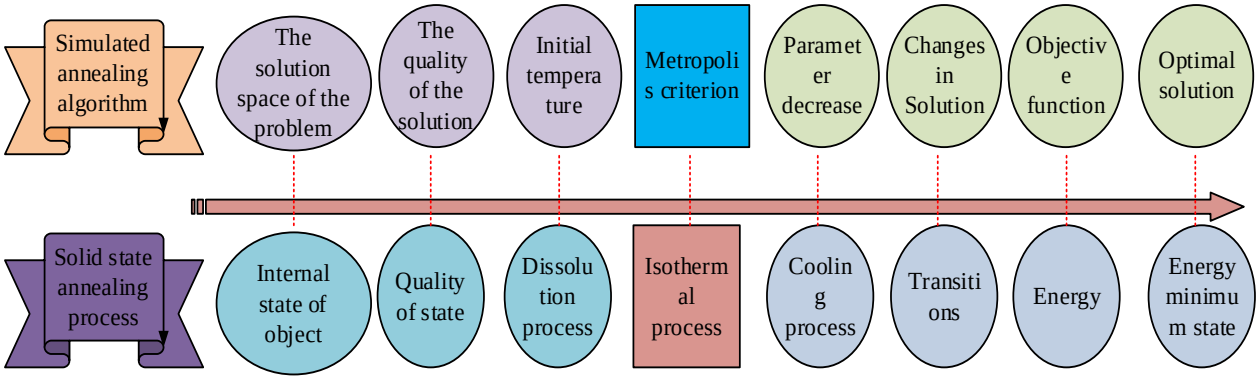


Fig. 3. The correspondence between simulated annealing algorithm and annealing process.

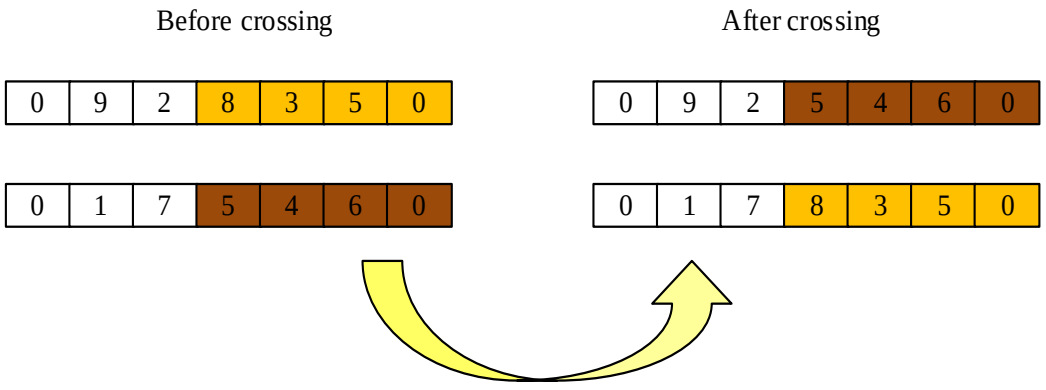


Fig. 4. Single point crossing effect before and after.

After applying crossover operations to genetic algorithms, the study improves the two-point variability theory by dividing variability into three stages: swapping, inversion, and insertion, forming a new mutation operator that avoids the local minima. After inserting the mutation, the Metropolis criterion is further introduced to determine whether the superior and inferior solutions are accepted. Firstly, the model function is solved to obtain an optimized solution, which serves as the initial value for the SA algorithm. Then, the SA algorithm is used to complete the local search in the surrounding space of the initial solution, and the local optimal solution is obtained from it. Finally, the Metropolis criterion is used to determine whether the new solution is accepted or not. When the difference between the model function of the local optimal solution and the original model function is less than 0, it can be determined as an acceptable new solution. Otherwise, it needs to be determined through Eq. (12).

$$p = \exp\left(-\frac{C_{new} - C_{old}}{T}\right) \quad (12)$$

In Eq. (12), T represents the current temperature. p represents the probability of the new solution being accepted. C_{new} represents the model function of the local optimal solution. C_{old} represents the original model function. In the probability of accepting a solution, a random number within the range of 0 to 1 is generated for comparison. When the random number is greater than p, the acceptable new solution can be retained and used as the solution of the model function. Otherwise, the new solution needs to be accepted, and the above steps need to be repeated continuously to update the annealing temperature. When the improved SA algorithm reaches the maximum number of iterations or continuously inputs the same result, the entire operation ends. Otherwise, it needs to be repeated continuously until the best result is obtained. Therefore, the operational flow of the improved SA algorithm can be presented in Fig. 5.

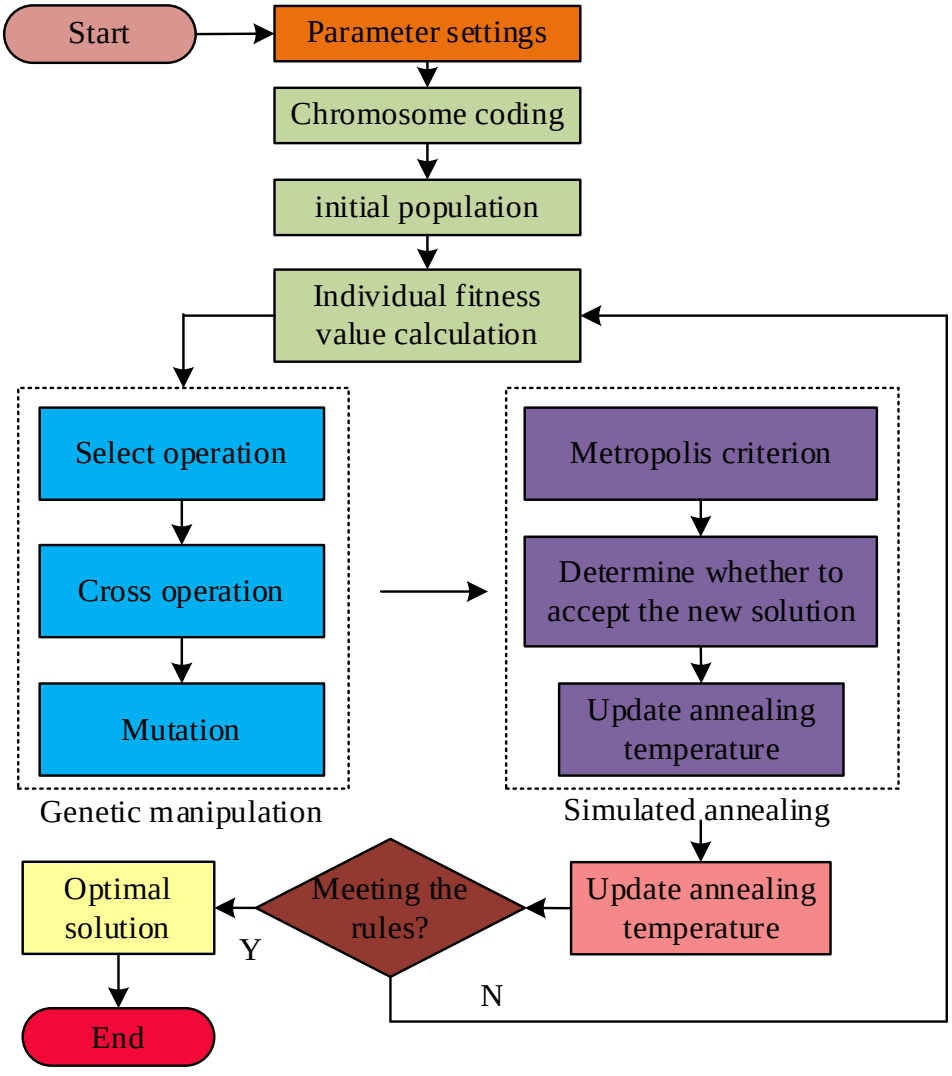


Fig. 5. The calculation process of the improved simulated annealing algorithm.

III. RESULTS

To verify the effectiveness of the proposed method for analyzing the environmental and economic benefits of urban construction projects, the study first validates the performance of the improved SA algorithm. Then, environmental and economic benefit analysis is conducted by combining cross efficiency DEA.

A. *Performance Testing of Improved Simulated Annealing Algorithm*

The performance testing of the improved SA algorithm mainly relies on Matlab 2020a simulation software, with Windows 11 operating system, 128GB of memory, and Intel Core i5-9300H central processor. During the testing process, traditional SA algorithm and SA-PSO algorithm are selected for comparison. The selected comparison indicators include fitness and F1 value, with an iteration of 2000 times. The dataset is CEC2022, which is a standard test function used to evaluate and compare the performance of optimization algorithms. This dataset can fully examine the search ability and robustness of optimization algorithms. The fitness and F1 value changes of the three methods are shown in Fig. 6.

From Fig. 6(a), during the variation of fitness values, the maximum fitness value of the improved SA was 0.86 and the minimum value was 0.21. It tended to stabilize after 480 iterations. The minimum fitness values of SA-PSO and traditional SA algorithms were 0.28 and 0.37, respectively, and both tended to stabilize after 1000 iterations, with a significantly slower convergence speed. According to Fig. 6(b), in the 5 experiments numbered A, B, C, D, and E, the maximum F1 value of the improved SA algorithm was 0.988 and the minimum value was 0.955. The average values of SA-PSO and traditional SA algorithms were 0.903 and 0.878, respectively. This proves that the improved SA algorithm has better overall performance. To further validate the performance of the improved SA algorithm, tests were conducted on two different datasets, including the Wine and Iris datasets in the UCI machine learning library. The Iris dataset is relatively small and simple, while the Wine dataset is relatively complex, and the relationships between features may be more complex, which can affect the convergence speed and stability of the algorithm. The proposed improved SA algorithm is more suitable for datasets with complex feature relationships or higher dimensions, as it can better explore and utilize the characteristics of the search space. The changes in fitness obtained are shown in Fig. 7.

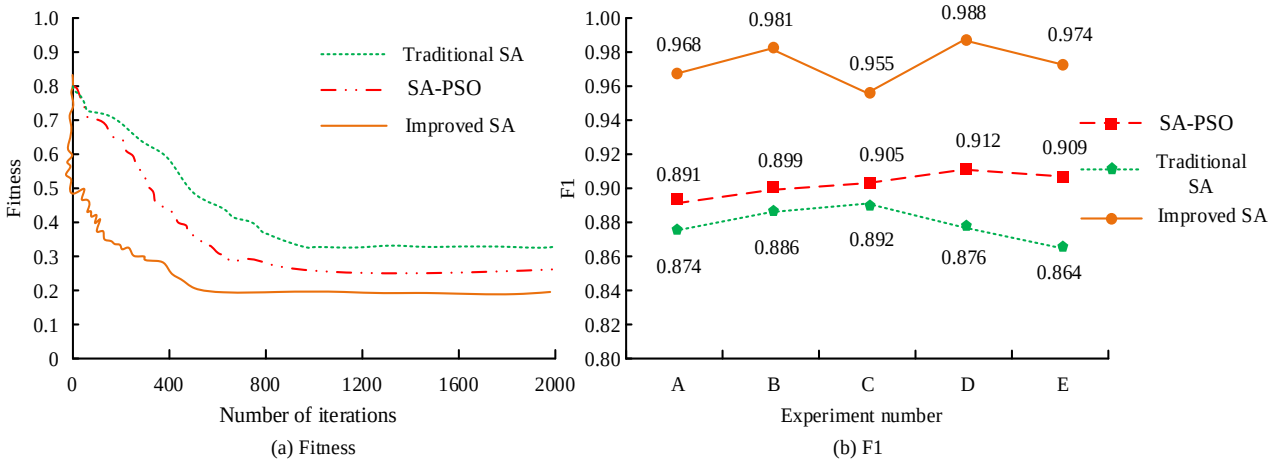


Fig. 6. The fitness and F1 value changes of three methods.

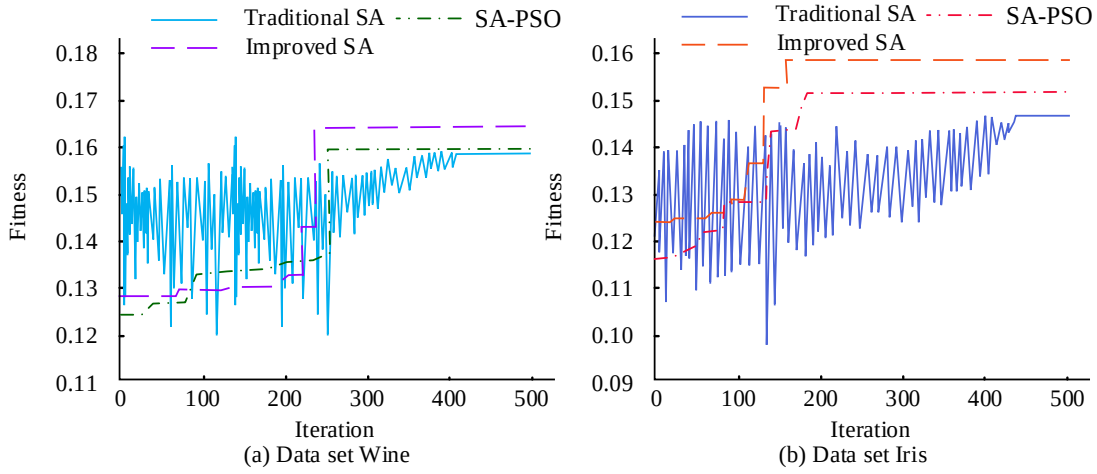


Fig. 7. Comparison of fitness values in different datasets.

From Fig. 7(a), the convergence curve of the traditional SA algorithm started with significant fluctuations and instability, and then decreased as the number of iterations increased. The algorithm eventually stabilized at 410 iterations, the fitness value was approximately 0.158. The convergence curve of the improved SA algorithm was smoother, rising in a ladder shape, and tended to stabilize at around 0.166 after 256 iterations. From Fig. 7(b), the convergence curve of the SA-PSO algorithm was relatively stable, but the improved SA algorithm still showed a ladder shape increase with a fitness value of about 0.157, indicating relatively better performance. The study continues to test the running time of the three algorithms. The results are shown in Fig. 8.

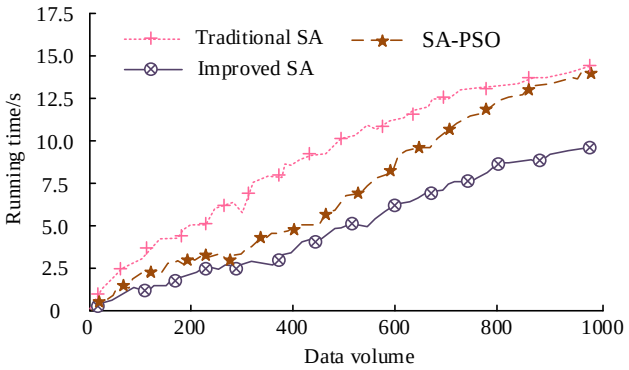


Fig. 8. The running time test results of three algorithms.

From Fig. 8, in terms of running time, with the increase of data volume, both the traditional SA algorithm and SA-PSO algorithm had a relatively fast increase in running time, with the highest being 14.2s and 14.6s, respectively. The running time curve of the improved SA algorithm always lowered than the other two algorithms, and the highest was only 9.7s, indicating significantly better running efficiency. Finally, the initial value dependency of the improved SA algorithm is validated, as shown in Table II.

From Table II, as for the improved SA algorithm, the standard deviation of the optimal value after 100 iterations was 0.00734, and the standard deviations of the optimization time and optimization iteration were 12.22352 and 3.55342, respectively. The standard deviations are relatively small, especially for the optimal value. This indicates that the stability of the optimization time and results is good, further proving the high robustness of the algorithm.

B. Empirical Analysis of Environmental and Economic Benefits of Urban Construction Projects

In order to analyze the effectiveness of the proposed method in analyzing the environmental and economic benefits of urban construction projects, three construction projects are selected from three provincial capitals in China, namely Project A, Project B, and Project C, with corresponding total construction areas of 21511m^2, 238500m^2, and 341000m^2, respectively. Firstly, the cross efficiency DEA model is used to solve the problem. The results are shown in Table III.

From Table III, the cross efficiency DEA model was effective for all three selected projects, achieving technical and scale efficiency in terms of cost and benefit, and having a high ecological efficiency. Both environmental and economic benefits were high. Based on this solution result, the improved SA algorithm is used to optimize the environment and economic benefits. The results are shown in Fig. 9.

From Fig. 9(a), in terms of environmental benefits, all three projects were improved to varying degrees, with the highest reaching 120000 yuan. From Fig. 9(b), after optimization, the economic benefits were significantly improved, with three projects increasing by 600000 yuan, 800000 yuan, and 850000 yuan respectively. From this, the proposed method can effectively improve the environmental and economic benefits of urban construction projects, and provide guidance for the implementation and later development of urban construction projects.

TABLE II. INITIAL VALUE DEPENDENCY VERIFICATION RESULTS

Solution results	Solution results	Optimization time	Optimal iterations
1	0.027416	82.78755	15
2	0.023774	75.24132	21
3	0.017184	88.46139	18
4	0.024701	93.57563	20
...	...	...	...
100	0.038525	81.86583	20
Standard deviation	0.00734	12.22352	3.55342

TABLE III. THE SOLUTION RESULTS OF THE CROSS EFFICIENCY DEA MODEL FOR THREE PROJECTS

Project	Insufficient output		Allocate redundancy		Ecological efficiency
	Economic benefits	*Environmental benefit*	*Indirect costs*	*Technical cost*	
Project A	0	0	0	0	1
Project B	0	0	0	2147.34	1
Project C	0	2.47	6.84	0	0.9987

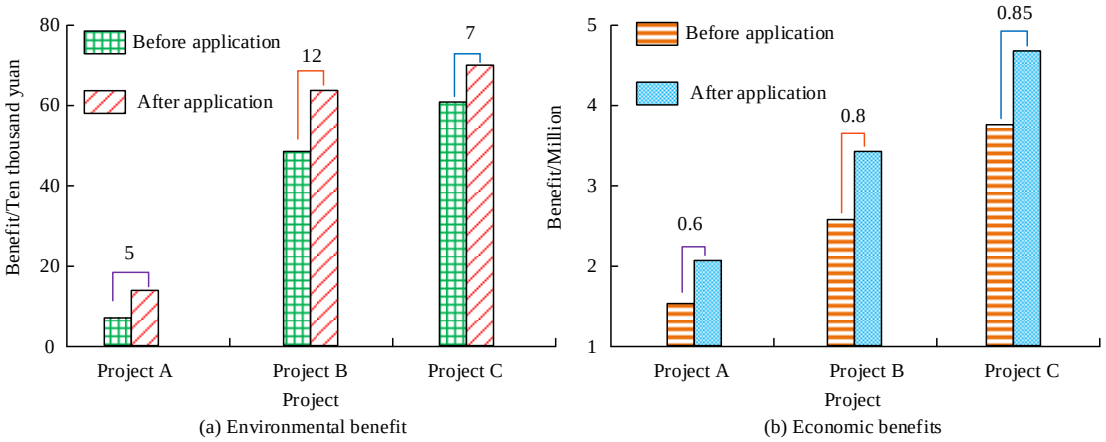


Fig. 9. Comparison of environmental and economic benefits before and after optimization.

IV. Discussion

To accurately analyze the economic and environmental benefits of urban construction projects, a cross efficiency DEA model was studied and designed, and a hybrid mutation operator and genetic algorithm were proposed to improve the SA algorithm. Finally, performance testing and practical verification were conducted. The results show that in the CEC2022 dataset, the maximum F1 value of the improved SA algorithm is 0.988 and the minimum value is 0.955, indicating better overall performance. The test results on different Wine and Iris datasets show that the convergence curve of the improved SA algorithm is smoother, and the fitness value ultimately stabilizes at 0.166. At the same time, the fitness curve of the improved SA algorithm shows a stepwise increase, resulting in better performance. In terms of running time, with the increase of data volume, the maximum running time of the improved SA algorithm is only 9.7 seconds, and the running efficiency is significantly better than the other two methods. Liu et al. introduced the SA algorithm into the adaptive particle swarm optimization algorithm to address the multi-objective allocation problem of maximizing weapon attack effectiveness, and proposed an adaptive SA optimization algorithm. The results show that the improved SA algorithm exhibits good performance in terms of convergence speed and global optimization ability, significantly better than a single SA algorithm, and can ensure maximum attack effectiveness [23]. The improved SA algorithm outperforms the basic SA algorithm in terms of performance. The difference is that in this study, a hybrid mutation operator and genetic algorithm were used to optimize the SA algorithm. The design of the hybrid mutation operator makes it easy to adjust the type and parameters of the mutation operator according to the characteristics of different construction projects, thereby optimizing the performance of the algorithm in specific scenarios. The population mechanism of genetic algorithm enables multiple solutions to be evaluated simultaneously, thereby improving the global optimization ability of the algorithm. However, Liu S et al. used adaptive particle swarm optimization algorithm for optimization, which lacks targeted optimization ability in specific scenarios. In addition, the applicability of the cross efficiency DEA model has been demonstrated in solving urban construction projects. After optimizing using the proposed method, the economic and environmental benefits of the three projects were significantly improved, with a maximum increase of 850000 yuan and 120000 yuan respectively, once again verifying the comprehensiveness of the proposed method in considering environmental and economic benefits and providing more comprehensive information. SoltanifarM et al. used a cross efficiency DEA model to rank decision units. The results show that this method can effectively consider the mutual influence between decision units when evaluating efficiency, thereby providing a more accurate and fair efficiency ranking [24]. By calculating cross efficiency, different decision-making units can not only compare their own inputs and outputs, but also use the efficiency of other decision-making units as a benchmark. This relative evaluation method makes the results more reliable. This once again confirms the effectiveness of the cross efficiency DEA model. The research provides powerful tools for decision support in actual urban construction projects, with broad application potential, and can promote sustainable urban construction in the future.

The improved SA algorithm and cross efficiency DEA model proposed in the study have achieved good performance in the economic and environmental benefit assessment of urban construction projects, but there are still some limitations. Firstly, the parameter selection and adjustment of the model depend on specific datasets, and the lack of automated parameter tuning mechanisms may affect its universality. Secondly, the model did not consider the impact of dynamic environmental changes on the efficiency of decision-making units, which may limit its applicability in practical applications. Future research can focus on enhancing the adaptive capability of models, introducing more complex multi-objective optimization frameworks, and exploring the application of new technologies such as deep learning in the optimization process to improve the robustness and real-time decision-making ability of algorithms.

V. Conclusion

A cross efficiency DEA model was studied and designed for the economic and environmental benefit analysis of urban construction projects, and an improved SA algorithm was proposed. The results showed that in the five experiments numbered A, B, C, D, and E, the maximum F1 value of the

improved SA algorithm was 0.988 and the minimum value was 0.955. The average values of SA-PSO and traditional SA algorithms are 0.903 and 0.878, respectively. This proves that the improved SA algorithm has better overall performance. In terms of running time, with the increase of data volume, the running time of both traditional SA algorithm and SA-PSO algorithm has grown rapidly, with the highest being 14.2s and 14.6s respectively. The running time curve of the improved SA algorithm has always been lower than the other two algorithms, and the highest is only 9.7s, indicating significantly better running efficiency. In addition, the applicability of the cross efficiency DEA model has been demonstrated in solving urban construction projects. After optimizing using the proposed method, the economic and environmental benefits of the three projects have significantly improved, with a maximum increase of 850000 yuan and 120000 yuan, respectively. The method proposed by the research institute can effectively consider the environmental and economic benefits in urban construction projects. In practical applications, by optimizing the improved SA algorithm, research can quickly identify projects with excellent performance, promote the rational allocation and use of resources, and enhance the overall sustainability of the project. Further research can explore the applicability of improved algorithms in other fields and promote their application in a wider range of decision support systems.

ACKNOWLEDGMENT

The research is supported by Science and Technology Research Project of Chongqing Municipal Education Commission, “Research on flexible dispatching mechanism of urban emergency materials under the COVID-19 epidemic - based on the perspective of the elderly” (KJQN202202505).

REFERENCES

[1] Opoku D G J, Agyekum K, Ayarkwa J. Drivers of environmental sustainability of construction projects: a thematic analysis of verbatim comments from built environment consultants. International Journal of Construction Management, 2022, 22(6): 1033-1041.

[2] Usman A M, Abdullah M K. An Assessment of Building Energy Consumption Characteristics Using Analytical Energy and Carbon Footprint Assessment Model. Green and Low-Carbon Economy, 2023, 1(1): 28-40.

[3] Joy Otibhor Olurin, Joachim Osheyor Gidiagba, Vincent Ebhohime Ehiaguina, Tina Chinyere Ndiwe, Gabriel Gbenga Ojo, Oluwaseun Ayo Ogunjobi. Safety, Quality Control, And Sustainability in Construction: Exploring the Nexus – A Review. Engineering Heritage Journal. 2023; 7(1): 72-93.

[4] Mabrouka Shahat Younis, Elfargani. The Benefits of Artificial Intelligence in Construction Projects. Acta Informatica Malaysia. 2022; 6(2): 47-51.

[5] Mohammad Allouzi, Mohammad Aljaafreh. Applied Ai in Neom Construction Projects: The Potential Impact of Ai in Enhancing Projects Success. Acta Informatica Malaysia. 2024; 8(1): 32-44.

[6] Mabrouka Shahat Younis, Elfargani. Assessing Construction Automation and Robotics in The Sustainability Sense. Engineering Heritage Journal. 2021; 5(2): 73-77.

[7] Wang Z, Tian J, Feng K. Optimal allocation of regional water resources based on simulated annealing particle swarm optimization algorithm. Energy Reports, 2022, 8: 9119-9126.

[8] Zhang Z, Wang L, Song X, Huang, H., & Yin, Y. Improved genetic-simulated annealing algorithm for seru loading problem with downward substitution under stochastic environment. Journal of the operational research society, 2022, 73(8): 1800-1811.

[9] Gabi D, Dankolo N M, Muslim A A, slim, A. A.. Dynamic scheduling of heterogeneous resources across mobile edge-cloud continuum using fruit fly-based simulated annealing optimization scheme. Neural Computing and Applications, 2022, 34(16): 14085-14105.

[10] Yu C, Chen M, Cheng K, Zhao, X., Ma, C., Kuang, F., & Chen, H. SGOA: annealing-behaved grasshopper optimizer for global tasks. Engineering with Computers, 2022, 38(5): 3761-3788.

[11] Liu J, Han C. Design and optimization of heat extraction section in energy tunnel using simulated annealing algorithm. Renewable Energy, 2023, 213: 218-232.

[12] Ch L K, Kamboj V K, Bath S K. Hybridizing slime mould algorithm with simulated annealing algorithm: a hybridized statistical approach for numerical and engineering design problems. Complex & intelligent systems, 2023, 9(2): 1525-1582.

[13] Camanho A S, Silva M C, Piran F S, & Lacerda, D. P. A literature review of economic efficiency assessments using Data Envelopment Analysis. European Journal of Operational Research, 2024, 315(1): 1-18.

[14] Zhu J. DEA under big data: Data enabled analytics and network data envelopment analysis. Annals of Operations Research, 2022, 309(2): 761-783.

[15] Cappa F, Franco S, Rosso F. Citizens and cities: Leveraging citizen science and big data for sustainable urban development. Business Strategy and the Environment, 2022, 31(2): 648-667.

[16] Chen S, Mao H, Sun J. Low-carbon city construction and corporate carbon reduction performance: evidence from a quasi-natural experiment in China. Journal of Business Ethics, 2022, 180(1): 125-143.

[17] Omopariola E D, Olanrewaju O I, Albert I, Oke, A. E., & Ibiyemi, S. B. Sustainable construction in the Nigerian construction industry: unsustainable practices, barriers and strategies. Journal of Engineering, Design and Technology, 2024, 22(4): 1158-1184.

[18] Sinuany-Stern Z. Foundations of operations research: From linear programming to data envelopment analysis. European Journal of Operational Research, 2023, 306(3): 1069-1080.

[19] Emrouznejad A, Marra M, Yang G, & Michali, M. Eco-efficiency considering NetZero and data envelopment analysis: a critical literature review. IMA journal of management mathematics, 2023, 34(4): 599-632.

[20] Amirteimoori A, Charles V, Mehdizadeh S. Stochastic data envelopment analysis in the presence of undesirable outputs. Journal of the Operational Research Society, 2023, 74(12): 2619-2632.

[21] Yildiz B S, Mehta P, Sait S M, Panagant, N., Kumar, S., & Yildiz, A. R. A new hybrid artificial hummingbird-simulated annealing algorithm to solve constrained mechanical engineering problems. Materials Testing, 2022, 64(7): 1043-1050.

[22] Liu C, Zhang F, Zhang H, Shi, Z., & Zhu, H. Optimization of assembly sequence of building components based on simulated annealing genetic algorithm. Alexandria Engineering Journal, 2023, 62: 257-268.

[23] Liu S, Liu W, Huang F, Yin, Y., Yan, B., Zhang, T. Multitarget allocation strategy based on adaptive SA-PSO algorithm. The Aeronautical Journal, 2022, 126(1300): 1069-1081.

[24] Soltanifar M, Sharafi H. A modified DEA cross efficiency method with negative data and its application in supplier selection. Journal of combinatorial optimization, 2022, 43(1): 265-296.

Improving the Accuracy of Chili Leaf Disease Classification with ResNet and Fine-Tuning Strategy

Sayuti Rahman[1]*, Rahmat Arief Setyadi[2], Asmah Indrawati[3], Arnes Sembiring[4], Muhammad Zen[5]
Department of Informatics Engineering-Faculty of Engineering, Universitas Medan Area, Indonesia[1, 2, 4]
Department of Agrotechnology-Faculty of Agriculture, Universitas Medan Area, Indonesia[3]
Sains dan Teknologi-Sistem Komputer, Universitas Pembangunan Panca Budi, Medan, Indonesia[5]

***Abstract*—Lack of diseases detection in plants frequently results in the spread of diseases that are difficult to treat and expensive. Rapid diseases recognition enables farmers to control the diseases with appropriate treatment. This study aims to support chili farmers in identifying chili plant diseases based on leaf images. This work presents a CNN design based on several existing CNN architectures that have been fine-tuned to achieve the highest possible accuracy. The study found that the ResNet101 model with the Tanh activation function, SGD optimizer, and Reduced Learning Rate (ReduceLR) schedule, achieved a peak classification accuracy of 99.53%. This significant improvement demonstrates the potential of using advanced CNN techniques and fine-tuning strategies to enhance model accuracy in agricultural applications. The implications of this study extend to the field of precision agriculture, suggesting that the proposed model can be integrated into smart farming systems to improve the timely and efficient control of chili leaf diseases. Such advancements not only enhance crop yields but also contribute to sustainable agricultural practices and the economic stability of chili farmers.**

***Keywords*—Chili leaf classification; convolutional neural network; ResNet10; fine-tuning; precision agriculture**

I. Introduction

Chili has been a flagship commodity in Indonesia's horticultural subsector, earning a solid global reputation. This finding emphasizes chili's enormous potential as a significant export commodity for Indonesia. The unique characteristics and high quality of Indonesian chili have earned global recognition, opening up significant opportunities for export growth and positively impacting the national economy [1], [2]. Chili yields a variety of products, such as fresh chili, dried chili, chili sauce, and chili oil. Dried chili is used as a spice, chili sauce is available in a variety of flavors, and chili oil, produced from chili, is utilized in the food, pharmaceutical, and cosmetic industries [3], [4].

Indonesian chili farmers face significant challenges that could affect production, welfare, and the long-term viability of their farming operations. These problems include environmental and weather fluctuations, which can have an impact on the health and growth of chili plants. In addition to changes in temperature and weather patterns, insect infestations such as thrips, caterpillars, and whiteflies, as well as diseases such as bacterial wilt and anthracnose, pose significant risks that require appropriate management measures [5]–[7]. Pest and disease attacks on chili plants are frequently detected based on the condition or appearance of the leaves, which serve as a food processing area for the plant [8] and poor chili plant development is a direct result of diseased leaves. Monitoring chili infections involves significant effort, and precise agricultural techniques are required to address this issue.

Several prior studies utilized computer vision to classify chili leaves in a variety of ways. Computer vision, which uses cameras and computational power to analyze and understand acquired images, can reliably classify diseases from chili leaves. Currently, the convolutional neural network (CNN) is the most used algorithm for computer vision problems. Transfer learning [9], fine-tuning [10], pruning [11], and new architectures [12] all have the potential to increase CNN performance and achieve high accuracy.

The aim of this study is to solve chili farmers' problems by developing a more efficient and accurate chili leaf disease detection application system utilizing CNN. The findings of this study show that by fine-tuning and optimizing the ResNet101 architecture, the accuracy of chili leaf disease classification may be greatly increased, outperforming previously classification approaches. The increased CNN accuracy attained in this work is intended to help establish a more effective and sustainable precision agriculture system. The study contributes the following:

- Analyzing how activation functions, optimization functions, and learning rate schedules improve CNN accuracy.
- Developing a fine-tuning model to increase CNN accuracy in Chili Leaf Disease Classification.
- Developing a more accurate CNN model suitable for smart farming systems.

II. Literature Review

Chili leaf classification research has advanced alongside developments in machine learning and computer vision technology. CNN has emerged as the preferred method for this purpose because of its ability to extract information from images automatically and accurately. Several methods and approaches have been investigated to increase classification accuracy, with a focus on improving network architecture and using data augmentation. Previous research has demonstrated that selecting of model and preprocessing procedures has a substantial impact on the overall system performance. Table I highlights many studies on chili leaf classification.

*Corresponding Author.

TABLE I. RELATED STUDIES

Metode	Number of Classes	Accuracy
VGGNet [13]	3	97.00%
SVM+RNN [14]	5	92.10%
GLCM+KNN [15]	2	94.00%
Fine Tuning ShuffleNet [16]	2	99.30%
Inception V3 [17]	4	93.00%

The more classes utilized in a classification system, the more diseases that affect chili leaves may be detected and classified. Thus, a more advanced and accurate system can offer more precise information for plant disease control. In this study, five classes were used to classify chili leaf diseases. According to previous studies, the accuracy achieved for the five classes was only 92.10%, showing that there is still space for development and greater accuracy.

This research focuses on enhancing the accuracy of CNNs for chili leaf disease classification by exploring various CNN architectures and optimization techniques. Techniques such as transfer learning and fine-tuning pre-trained models are expected to significantly improve model performance. Additionally, the use of more diverse data augmentation is anticipated to help mitigate overfitting and improve the model's generalization ability to new data. This study aims to contribute to the field of precision agriculture by providing more reliable and efficient solutions for plant disease detection.

III. RESEARCH METHODS

This study was conducted to evaluate several aspects of the implementation and refinement of CNN architectures, with a focus on ResNet. To achieve this goal, a series of experiments were performed, including comparing accuracy on both augmented and non-augmented data, comparing the number of layers in ResNet, and fine-tuning parameters such as activation functions, optimizers, and learning rate schedules. The research workflow began with an evaluation of the effect of data augmentation and continued with a comparison of the performance of CNN models on augmented data. Subsequently, the study investigated the effects of different ResNet layer numbers on accuracy. Experiments with different optimization functions, types of optimizers, and learning rate schedules were also part of the fine-tuning process. Finally, the best results were compared with previous studies to identify improvements achieved.

A. Fine-Tuning Strategy

Fine-tuning is performed to determine the most accurate CNN model. In this work, fine-tuning is performed on an existing CNN that has been tested and produces the highest accuracy, namely the Resnet architecture. Fine-tuning is completed in three stages: activation function tuning, optimizer tuning, and learning rate schedule. The CNN with the best activation function is saved and then fine-tuned by the optimizer. The most accurate activation function and optimizer are reported, and the learning rate schedule is adjusted accordingly. As a result, this study creates the most accurate CNN model. The fine-tuning method is illustrated in Fig. 1.

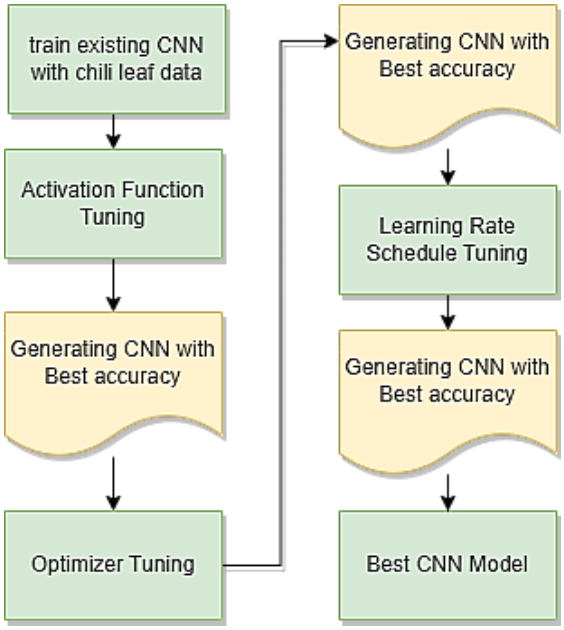


Fig. 1. Fine tuning strategy.

B. Dataset and Augmentation

The dataset used in this research consists of images of diseases affecting red chili leaves, sourced from Mendeley Data, comprising five disease classes [18]. The dataset is divided into two types: augmented and non-augmented. The dataset originally contained 531 images before augmentation, and after augmentation, the dataset expanded to 2,128 images. The details of the augmented dataset are shown in Table II below:

TABLE II. AUGMENTED DATASET

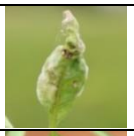
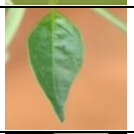
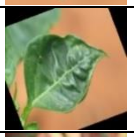
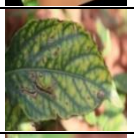
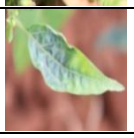
Class Name	Image	Training	Validation	Total
Powdery Mildew		486	122	608
Healthy Leaf		221	55	276
Murda Complex (mites, thrips)		342	86	428
Leaf Spot (Cercospora)		326	82	408
Nutrient Deficiency		327	81	408
Total		1702	426	2128

Table II provides information on the distribution of the augmented dataset used in this research to identify diseases in red chili leaves. This dataset consists of five different disease classes: Powdery Mildew, Healthy Leaf, Murda Complex (mites, thrips), Leaf Spot (Cercospora), and Nutrient Deficiency. Each class contains a number of images that have been divided into subsets for training and validation. The total number of images in the dataset after augmentation is 2,128, which includes 1,702 images for training and 426 images for validation. The original images that have been augmented are shown in Fig. 2.

Fig. 2. Data augmentation.

C. Activation Function

The selection of an appropriate activation function in Convolutional Neural Networks (CNNs) is crucial for improving model accuracy, particularly in classification tasks. Activation functions like ReLU, which are popular for their simplicity and effectiveness, help the network learn complex non-linear representations. However, in some cases, ReLU can lead to issues such as dead neurons, which can be mitigated by variants like Leaky ReLU or ELU that introduce a slope or non-linearity on the negative side. Other functions, such as Tanh and Sigmoid, also play important roles, particularly in regulating the output within specific ranges. Tanh maps inputs to values between -1 and 1, while Sigmoid maps inputs to values between 0 and 1. For multi-class classification, Softmax is used to convert outputs into probabilities, enabling more accurate predictions. By understanding the characteristics of these activation functions, we can optimize CNN architectures to achieve the best performance in various classification applications.

The linear activation function returns the input without any modification. It is typically used in the final layer of regression models. The equation for the linear activation function is presented in Eq. (1) [19]:

$$f(x) = a.x \tag{1}$$

ReLU (Rectified Linear Unit) is a highly popular non-linear activation function that returns the input directly if it is positive, and zero if it is negative. The equation for ReLU is shown in Eq. (2) [20].

$$f(x) = \max(0, x) \tag{2}$$

ELU (Exponential Linear Unit) is an activation function similar to ReLU but differs in how it handles negative inputs. ELU introduces non-linearity on the negative side to address the dead neuron problem associated with ReLU[21]. The equation for ELU is presented in Eq. (3).

$$f(x) = \begin{cases} x, & if\ x > 0 \\ \alpha(\exp(x) - 1), & if\ x \leq 0 \end{cases} \tag{3}$$

Leaky ReLU is a variation of ReLU that allows a small gradient when the input is negative [22]. The equation for Leaky ReLU is shown in Eq. (4).

$$f(x) = \begin{cases} x, & if\ x > 0 \\ \alpha x, & if\ x \leq 0 \end{cases} \tag{4}$$

Tanh (Hyperbolic Tangent) is an activation function that maps the input to a value between -1 and 1. It is often used in hidden layers of neural networks [23]. The equation for Tanh is given in Eq. (5):

$$f(x) = \text{Tanh}(x) \tag{5}$$

The Sigmoid activation function maps the input to a value between 0 and 1, which is commonly used in binary classification models [24]. The equation for the Sigmoid function is presented in Eq. (6):

$$f(x) = \frac{1}{1+\exp(-x)} \tag{6}$$

Softmax is an activation function typically used in the output layer for multi-class classification problems [25]. It converts the input values into probabilities. The equation for Softmax is shown in Eq. (7):

$$f(x) = \frac{\exp(x_i)}{\sum_{j=1}^{N} \exp(x_j)} \tag{7}$$

IV. Results and Discussions

A. Accuracy on Augmented and Non-Augmented Datasets

This study utilized two types of chili image datasets: one with augmented data and one without. The augmentation process involved increasing the dataset by applying horizontal flips, vertical flips, and 30-degree rotations. During model training, several key settings influenced how the model learned from the data. One of these was the number of epochs, which represents the number of times the model processes the entire dataset—set to 30 in this case. The batch size, which refers to the number of data samples processed together, was set to 32, impacting the speed and efficiency of the model's learning process. Additionally, the learning rate, fixed at 0.001, determined the magnitude of adjustments made during training to enhance the model's predictions. The momentum, set at 0.9 in the SGDM optimizer, was used to accelerate the convergence of the model. Initial experiments compared the accuracy of the ResNet101 architecture, with the results summarized in Table III.

TABLE III. ACCURACY IMPROVEMENT WITH DATASET AUGMENTATION

No	Method	Training Accuracy	Validation Accuracy
1	Non-Augmented Dataset ResNet101ResNet101	98.35%	89.72%
2	Augmented Dataset ResNet101	97.65%	98.12%

In the experiments using the ResNet101 architecture, two scenarios were compared. In the first scenario, without dataset augmentation, the model achieved a training accuracy of 98.35% and a validation accuracy of 89.72%. In the second scenario, with dataset augmentation, the model reached a training accuracy of 97.65% and a validation accuracy of 98.12%. The effect of augmentation is evident in the unusual outcome where the validation accuracy exceeded the training accuracy. However, applying dataset augmentation led to a significant improvement in the model's performance on the validation set, making it a more advantageous approach.

B. Accuracy on Existing CNNs

Based on the previous testing, it was found that data augmentation successfully improved the accuracy of CNNs in classifying chili leaf diseases. Using the augmented dataset, we tested several existing CNN architectures to determine which one would be more suitable for this classification task. The existing CNNs tested included AlexNet, GoogleNet, VGGNet, and ResNet. Table IV presents the key performance metrics obtained, including training accuracy and validation accuracy.

TABLE IV. ACCURACY OF EXISTING CNNS

No	Method	Training Accuracy	Validation Accuracy
1	AlexNet	93.9%	95.77%
2	GoogleNet	95.12%	96.95%
3	VGGNet16	95.95%	96.95%
4	ResNet 101	97.64%	98.12%

The table above shows the training accuracy and validation accuracy for several convolutional neural network methods used in the experiments. AlexNet achieved a training accuracy of 93.89% and a validation accuracy of 95.77%. GoogleNet demonstrated a training accuracy of 95.12% and a validation accuracy of 96.95%. VGGNet16 reached a training accuracy of 95.95% and a validation accuracy of 96.95%. ResNet101 achieved the highest accuracy, with a training accuracy of 97.65% and a validation accuracy of 98.12%.

C. Accuracy Comparison on ResNet

The comparison of the performance of various layers within the ResNet architecture used in this study aims to evaluate the effectiveness of each ResNet layer in the developed classification model. The comparison results are presented in Table V.

Table V presents a comparison of the training and validation accuracy of various ResNet methods used in the study. The ResNet18 method achieved a training accuracy of 96.42% and a validation accuracy of 95.77%. ResNet34 showed an improvement with a training accuracy of 97.36% and a validation accuracy of 97.18%. The ResNet50 method achieved a training accuracy of 97.83% and a validation accuracy of 97.18%. ResNet101 reached a training accuracy of 97.65% and the highest validation accuracy among the methods, at 98.12%. Finally, the ResNet152 method achieved the highest training accuracy of 98.41%, but its validation accuracy reverted to 97.18%.

D. Fine-Tuning Activation Functions for ResNet101

In an effort to enhance the performance of the ResNet101 model in classifying chili leaf diseases, fine-tuning was performed on the activation functions. Activation functions play a crucial role in determining the output of each layer in a neural network. In this section, various activation functions, such as ReLU, ELU, Leaky ReLU, Tanh, Sigmoid, and Softmax, were trained over 30 epochs. The impact of each activation function on the model's performance was then evaluated. The accuracy results for each activation function are presented in Table VI.

The model was trained using various activation methods, yielding different training and validation accuracies. The Linear method achieved a training accuracy of 97.65% and a validation accuracy of 98.12%. The ReLU method resulted in a training accuracy of 97.30% and a validation accuracy of 97.65%, while the ELU method produced a training accuracy of 97.71% and a validation accuracy of 96.95%. The Leaky ReLU method showed a training accuracy of 98.24% and a validation accuracy of 97.65%. The Tanh method achieved a training accuracy of 97.24% and the highest validation accuracy among all methods, at 99.06%. The Sigmoid method resulted in a training accuracy of 94.65% and a validation accuracy of 95.77%, while the Softmax method achieved a training accuracy of 93.89% and a validation accuracy of 96.48%. The validation results for each epoch are illustrated in Fig. 3.

TABLE V. COMPARISON OF RESNET LAYERS

No.	Method	Training Accuracy	Validation Accuracy
1	ResNet18	96.42%	95.77%
2	ResNet34	97.36%	97.18%
3	ResNet50	97.83%	97.18%
4	ResNet101	97.65%	98.12%
5	ResNet152	98.41%	97.18%

TABLE VI. ACCURACY WITH DIFFERENT ACTIVATION FUNCTIONS

No	Method	Training Accuracy	Validation Accuracy
1	Linear	97.65%	98.12%
2	ReLU	97.30%	97.65%
3	ELU	97.71%	96.95%
4	Leaky ReLU	98.24%	97.65%
5	Tanh	97.24%	99.06%
6	Sigmoid	94.65%	95.77%
7	Softmax	93.89%	96.47%

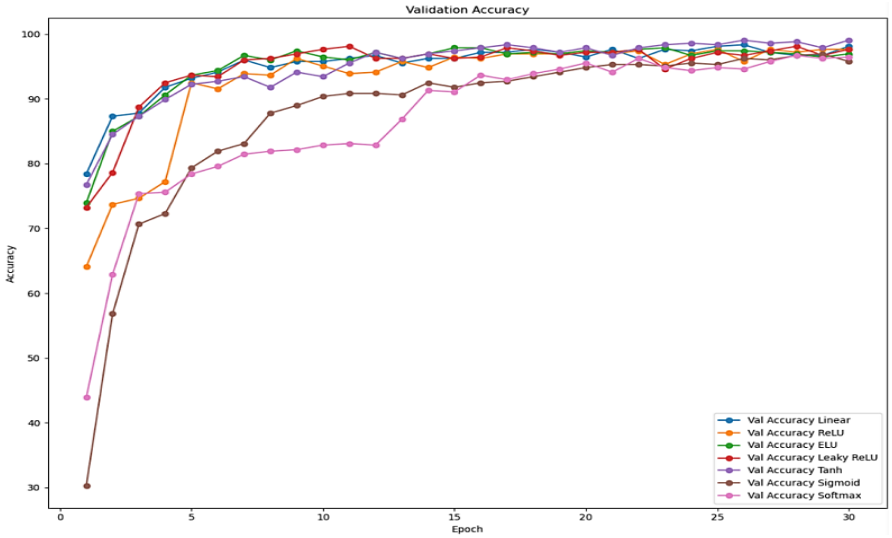


Fig. 3. Graph comparing validation accuracy with different activation functions.

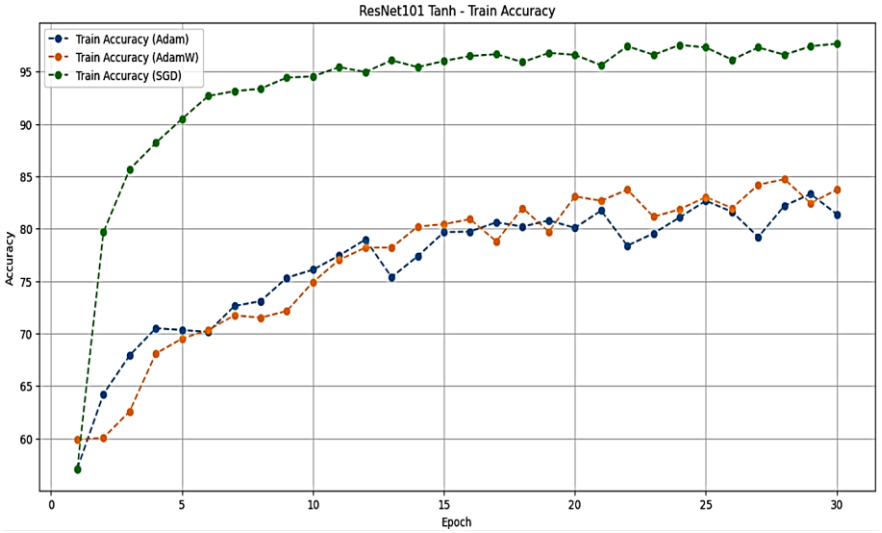


Fig. 4. Comparison of training accuracy using different optimizers.

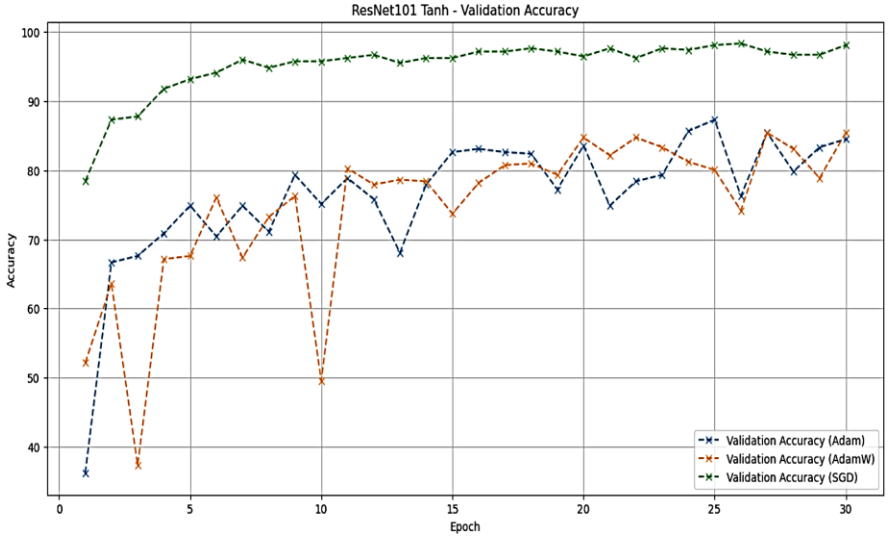


Fig. 5. Comparison of validation accuracy using different optimizers.

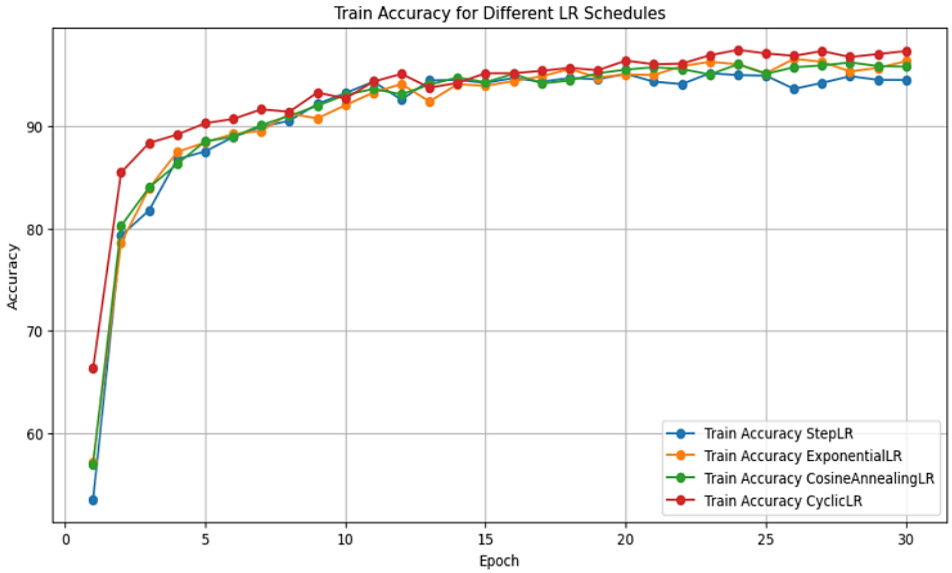


Fig. 6. Comparison of training accuracy across different learning rate schedules.

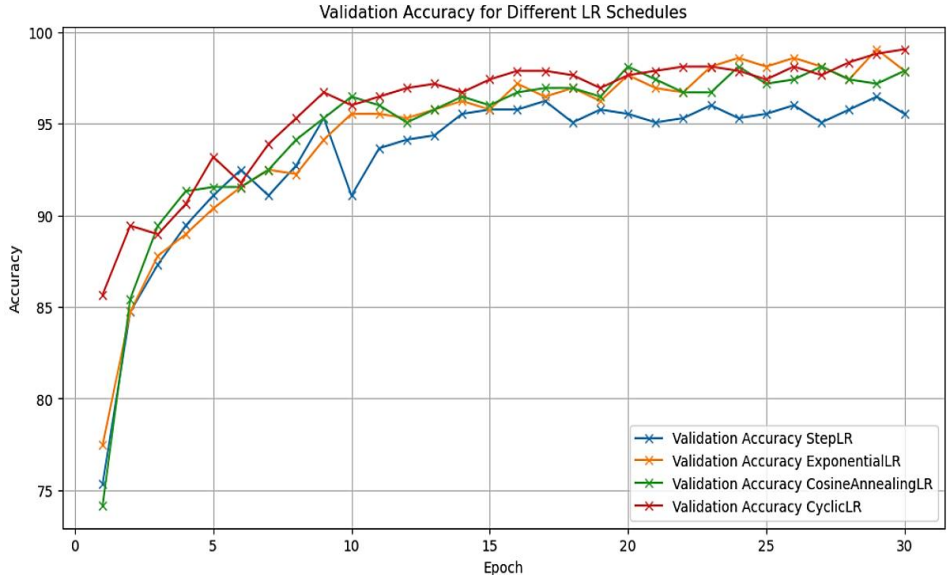


Fig. 7. Comparison of validation accuracy across different learning rate schedules.

E. Fine Tuning Optimizer ResNet101

Based on previous testing, ResNet with the Tanh activation function achieved the highest accuracy. To identify the optimal optimization function, several optimizers were tested while maintaining the Tanh activation function. The initial step involved selecting the appropriate optimizer to effectively handle the task of classifying chili leaf images. The optimizers tested included Adam, AdamW, and SGDM. The results of these tests are presented in Table VII.

Table VII shows the accuracy results of the three machine learning models tested. The different optimization functions yielded varying results. The SGDM optimizer demonstrated the best performance on both training and validation data, with accuracies of 97.24% and 99.06%, respectively. The AdamW model achieved a training accuracy of 84.72% and a validation accuracy of 85.45%, while the Adam model reached a training accuracy of 83.31% and a validation accuracy of 87.32%. Therefore, ResNet101 with the SGDM optimizer achieved the best accuracy in chili leaf classification. The accuracy comparisons for each epoch are presented in the Fig. 4 and 5.

TABLE VII. COMPARISON OF OPTIMIZERS

No	Method	Training Accuracy	Validation Accuracy
1	Adam	83.31%	87.32%
2	AdamW	84.72%	85.45%
3	**SGD**	**97.24%**	**99.06%**

F. Fine Tuning Learning Rate Schedule

In this study, five different learning rate schedules were evaluated to determine their impact on training and validation

accuracy. The results of this evaluation are presented in Table VIII below.

TABLE VIII. COMPARISON OF LEARNING RATE SCHEDULES

No	Function	Training Accuracy	Validation Accuracy
1	StepLR	95.1234	96.4789%
2	ExponentialLR	96.3572%	99.0610%
3	ConsineAnnealigLR	96.2397%	98.1221%
4	CyclicLR	97.3561%	99.0610%
5	ReduceLR	97.7673%	99.5305%

Table VIII provides a comparison of the training and validation accuracies for each learning rate schedule model tested. The StepLR model achieved a training accuracy of 95.12% and a validation accuracy of 96.48%. The ExponentialLR model, on the other hand, showed an improved training accuracy of 96.36% and a validation accuracy of 99.06%. The CosineAnnealingLR model attained a training accuracy of 96.24% and a validation accuracy of 98.12%. The CyclicLR model demonstrated further improvement with a training accuracy of 97.36% and a validation accuracy of 99.06%. Finally, the ReduceLR model exhibited the highest performance, achieving the highest training accuracy of 97.77% and the highest validation accuracy of 99.53%, making it the best-performing model among those tested.

The results indicate that different learning rate schedules can significantly affect the performance of the ResNet101 model. While StepLR provided a stable training process, models like ExponentialLR and CyclicLR demonstrated higher validation accuracies, suggesting they were better at generalizing to unseen data. The CosineAnnealingLR schedule showed a balance between training and validation accuracy, but the ReduceLR model outperformed all others by achieving both the highest training and validation accuracies. This suggests that dynamically reducing the learning rate as training progresses may be the most effective strategy for optimizing model performance in this context.

To visualize the impact of these learning rate schedules over the course of training, the accuracy for each epoch is plotted in the Fig. 6 and 7.

These figures illustrate how each learning rate schedule influenced the model's ability to learn and generalize from the data over time. As shown, the ReduceLR schedule not only maintained a high training accuracy but also led to the best validation accuracy, indicating its effectiveness in avoiding overfitting and improving model robustness.

G. Implementation

After completing the model evaluation phase on the test data, this research also implemented the model into an application. The application serves as a platform that facilitates users in classifying chili leaves to identify potential diseases. The classification page of the application is shown in Fig. 8.

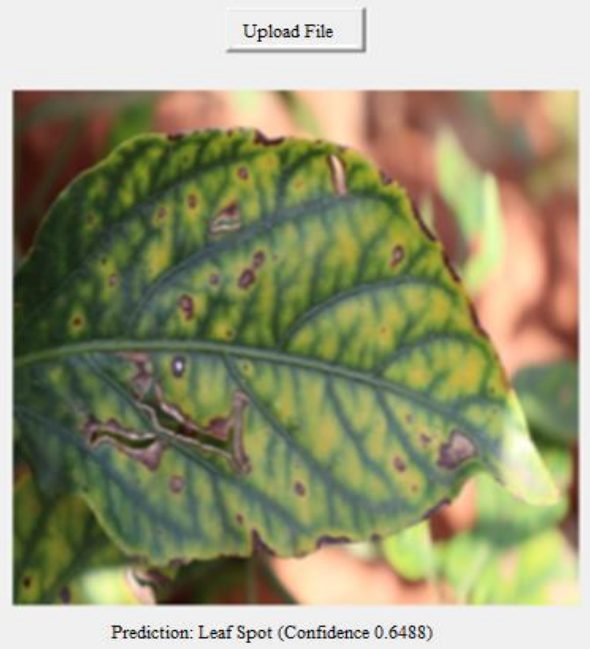


Fig. 8. Implementation of chili leaf disease classification.

Fig. 8 illustrates the classification process within the application. The image of the chili leaf, captured by a camera, is successfully classified by the ResNet101 model using the best-performing parameters. The application interface provides clear feedback to the user, indicating the identified disease, thus aiding in quick and accurate diagnosis.

H. Comparison with Previous Research

This study successfully classified chili leaf diseases with high accuracy, surpassing the accuracy achieved in previous studies. The highest accuracy was obtained with the classification of five chili leaf classes. A comparison of the accuracy from this study with that of previous research is presented in Table IX.

TABLE IX. ACCURACY COMPARISON WITH PREVIOUS RESEARCH

Methods	Number of Classes	Accuracy
VGGNet [13]	3	97.00%
SVM+RNN [14]	5	92.10%
GLCM+KNN [15]	2	94.00%
Fine Tuning ShuffleNet [16]	2	99.30%
Inception V3 [17]	4	93.00%
Xception [18]	5	79.56%
Resnet 101 (Best Parameter)	5	99.53%

According to Table IX, ResNet101 with the Tanh activation function, SGD optimizer, and ReduceLR learning rate schedule achieved an accuracy of 99.53%, accepted as the ResNet101 best parameter configuration. This best-parameter ResNet101 outperformed SVM+RNN [14], which used the same number of classes, with a 7.43% increase in accuracy. ResNet101 outperformed Xception [18], on the exactly the same dataset, with a 19.97% improvement. The best-parameter

ResNet101 also outperformed Fine Tuning ShuffleNe [16], which was only classify two chili leaf disease classifications.

V. CONCLUSION

This study successfully developed and refined a Convolutional Neural Network (CNN) model, especially ResNet101, for classifying chili leaf diseases with reliable accuracy. Through a series of experiments, including the application of multiple activation functions, optimization functions, and learning rate schedules, the model obtained an accuracy of 99.53% when diagnosing five different types of chili leaf diseases. This performance outperforms prior studies that showed the effectiveness of the ResNet101 model using the Tanh activation function, SGD optimization function, and ReduceLR learning rate schedule.

The findings demonstrated that carefully selecting and refining model parameters can greatly increase CNN accuracy in complex classification tasks. The proposed model not only outperformed previous CNN architectures such as VGGNet, Inception V3, and Xception, but it was also more effective in handling several disease classifications than other approaches such as SVM+RNN and Fine Tuning ShuffleNet.

Overall, implementing this model in a practical application-based system has the potential to help farmers and agricultural professionals in accurately diagnosing chili leaf diseases, resulting in improved disease management and crop yields. However, there are limitations to this study. Firstly, the model's performance has not been extensively tested on datasets collected from diverse environmental conditions, which may affect its generalizability in real-world scenarios. Secondly, the computational resources required for training and deploying the ResNet101 model may pose challenges for integration into low-cost, resource-constrained agricultural systems commonly found in rural areas. Further research might explore possibilities into refining the model and applying it to other agricultural fields, while addressing these limitations to enhance its practical applicability and efficiency.

ACKNOWLEDGMENT

This research was funded by the Directorate of Research, Technology, and Community Service (DRTPM), Directorate General of Higher Education, Ministry of Education and Culture of the Republic of Indonesia in 2024 under contract No. 103/E5/PG.02.00.PL/2024, with derivative contracts 020/LL1/AL.04.03/2024, and 52/P3MPI/01.3.2/VI/2024. The author would like to thank DRTPM and the Center for Research, Community Service, and International Publication (P3MPI) of Medan Area University for their support.

REFERENCES

[1] M. T. Sundari, D. Darsono, J. Sutrisno, and E. Antriyandarti, "Analysis of trade potential and factors influencing chili export in Indonesia," Open Agriculture, vol. 8, no. 1, p. 20220205, 2023.

[2] M. T. Sundari, D. Darsono, J. Sutrisno, and E. Antriyandarti, "Analysis of chili demand in Indonesia," in AIP Conference Proceedings, 2023, vol. 2583, no. 1.

[3] A. Benbouriche, Y. Benchikh, H. Boudries, and H. Guemghar-Haddadi, "The industrial by-product of chili paste: optimized carotenoids extraction," Algerian Journal of Environmental Science and Technology, vol. 7, no. 3, 2021.

[4] A. J. Fernando and S. Amaratunga, "Application of far infrared radiation for sun dried chili pepper (Capsicum annum L.): drying characteristics and color during roasting," Journal of the Science of Food and Agriculture, vol. 102, no. 9, pp. 3781–3787, 2022.

[5] P. R. Shingote et al., "An overview of chili leaf curl disease: Molecular mechanisms, impact, challenges, and disease management strategies in Indian subcontinent," Frontiers in Microbiology, vol. 13, p. 899512, 2022.

[6] M. Bulle, V. Sheri, M. Aileni, and B. Zhang, "Chloroplast Genome Engineering: A Plausible Approach to Combat Chili Thrips and Other Agronomic Insect Pests of Crops," Plants, vol. 12, no. 19, p. 3448, 2023.

[7] A. Sembiring, R. S. Basuki, R. Rosliani, and S. T. Rahayu, "Farmers' challenges on chili farming in the acid dry land: A case study from Pasir Madang-Bogor Regency, Indonesia," in E3S Web of Conferences, 2021, vol. 316, p. 3010.

[8] W. Setiawati et al., "GROWTH, AND YIELD CHARACTERISTICS AS WELL AS PESTS AND DISEASES SUSCEPTIBILITY OF CHILI PEPPER (Capsicum annuum L.) UNDER DIFFERENT PLANT DENSITIES AND PRUNING LEVELS.," Applied Ecology & Environmental Research, vol. 20, no. 1, 2022.

[9] A. W. Salehi et al., "A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope," Sustainability, vol. 15, no. 7, p. 5930, 2023.

[10] S. Rahman, M. Ramli, F. Arnia, R. Muharar, and A. Sembiring, "Performance analysis of mAlexnet by training option and activation function tuning on parking images," IOP Conference Series: Materials Science and Engineering, vol. 1087, no. 1, p. 012084, Feb. 2021, doi: 10.1088/1757-899X/1087/1/012084.

[11] S. Rahman et al., "Mini shufflenet for efficient parking space classification," in AIP Conference Proceedings, 2023, vol. 2480, no. 1, p. 030007, doi: 10.1063/5.0103430.

[12] S. Rahman, M. Ramli, F. Arnia, R. Muharar, M. Ikhwan, and S. Munzir, "Enhancement of convolutional neural network for urban environment parking space classification," Global Journal of Environmental Science and Management, vol. 8, no. 3, pp. 315–326, 2022, doi: https://doi.org/10.22034/gjesm.2022.03.02.

[13] A. F. K. Hasbollah, Z. M. Zin, N. Ibrahim, and R. F. R. Suleiman, "Green Chili Leaf Disease Detection Using Convolution Neural Networks," Journal of Green Engineering, vol. 10, pp. 13005–13019, 2020.

[14] N. N. Ahmad Loti, M. R. Mohd Noor, and S. W. Chang, "Integrated analysis of machine learning and deep learning in chili pest and disease identification," Journal of the Science of Food and Agriculture, vol. 101, no. 9, pp. 3582–3594, 2021, doi: 10.1002/jsfa.10987.

[15] A. Patil and K. Lad, "Feature Selection for Chili Leaf Disease Identification Using GLCM Algorithm," in IOT with Smart Systems: Proceedings of ICTIS 2021, Volume 2, 2022, pp. 359–365.

[16] C. J. Entuni, T. M. A. Zulcaffle, and K. H. Ping, "Classification of capsicum leaf disease from a complex cluster of leaves using an improved multiple layers ShuffleNet CNN model," International Journal of Advanced Technology and Engineering Exploration, vol. 10, no. 102, p. 515, 2023.

[17] Z. Gulzar, S. Chandu, and K. Ravi, "Classification and Analysis of Chili Plant Disease Detection Using Convolution Neural Networks," in International Conference on Image Processing and Capsule Networks, 2023, pp. 677–696.

[18] M. P. Aishwarya and A. P. Reddy, "Dataset of chili and onion plant leaf images for classification and detection," Data in Brief, vol. 54, p. 110524, 2024.

[19] S. Sharma, S. Sharma, and A. Athaiya, "Activation functions in neural networks," International Journal of Engineering Applied Sciences and Technology, vol. 4, no. 12, pp. 310–316, 2020.

[20] H.-S. Feng and C.-H. Yang, "PolyLU: A simple and robust polynomial-based linear unit activation function for deep learning," IEEE Access, 2023.

[21] Z. Feng, Z. Wang, K. Zheng, R. Li, Y. Zhao, and Y. Wang, "Enhancing deterministic prediction in unidirectional ocean waves using an Artificial Neural Network with exponential linear unit," Ocean Engineering, vol. 301, p. 117539, 2024.

[22] S. Padshetty and Ambika, "Leaky ReLU-ResNet for Plant Leaf Disease Detection: A Deep Learning Approach," Engineering Proceedings, vol. 59, no. 1, p. 39, 2023.

[23] S.-L. Shen, N. Zhang, A. Zhou, and Z.-Y. Yin, "Enhancement of neural networks with an alternative activation function tanhLU," Expert Systems with Applications, vol. 199, p. 117181, 2022.

[24] M. Mesran, S. R. Yahya, F. Nugroho, and A. P. Windarto, "Investigating the Impact of ReLU and Sigmoid Activation Functions on Animal Classification Using CNN Models," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 8, no. 1, pp. 111–118, 2024.

[25] T. Pearce, A. Brintrup, and J. Zhu, "Understanding softmax confidence and uncertainty," arXiv preprint arXiv:2106.04972, 2021.

Exploring the Impact of User Experience Elements on Virtual Reality for Emotion Regulation Through mVR-Real App

Irna Hamzah[1], Ely Salwana[2], Nilufar Baghaei[3]
Institute of Visual Informatics, Universiti Kebangsaan Malaysia (UKM), Bangi, Selangor, Malaysia[1, 2]
School of Electrical Engineering and Computer Science, The University of Queensland, St Lucia, Australia[3]

***Abstract*—Virtual reality, a rapidly advancing technological development, has significantly evolved over the past few years, offering immersive experiences through headsets, gloves, or controllers that engage users in dynamic and captivating environments. Its applications span various fields, including video games, healthcare, education, and training simulations. However, there remains a gap in utilizing virtual reality for adolescent emotion regulation. This study explores the mVR-REAL app, designed to enhance social-emotional learning in teenagers by improving emotion regulation. Social-emotional learning is an educational approach aimed at developing essential socio-emotional skills in adolescents. Sixteen participants engaged with the mVR-REAL app through virtual scene episodes using the Meta Quest 2 headset. The emotional responses and user feedback were measured using pre and post-test questionnaires with a Likert scale such as, strongly agree, agree, neutral, disagree and strongly disagree. A numerical value was assigned to each hypothetical option on the scale. Analysis conducted with SPSS software revealed statistically significant improvements in user experience following the use of mVR-REAL. The findings suggest that mVR-REAL has the potential to enhance user experience, evoke strong emotional responses, and foster greater engagement compared to traditional applications. These insights will inform future large-scale testing of mVR-REAL, emphasizing the importance of emotional design, as well as psychological and cultural factors, in the development of virtual reality apps for emotion regulation. However, time-related challenges were identified due to the restricted duration of the VR session, highlighting the need for further research and refinement in future virtual reality app development.**

Keywords*—Virtual reality; app; user experience; emotion regulation; social-emotional learning

I. Introduction

Virtual reality (VR) is an advanced technology that creates a realistic and all-encompassing experience by simulating various settings and activities. The term "virtual reality" refers to a computer-generated simulation of a three-dimensional setting [1] or experience [2] that enables the user to interact with the simulation in a manner that is convincingly real or physical by employing specialized hardware such as a headset, gloves, or sensors. This cutting-edge technology generates sensory stimuli, serving as a transformative tool for enhancing emotion regulation (ER) practices, offering unique immersive environments that facilitate emotional experiences and coping strategies.

The study was sponsored by Malaysia's Ministry of Education under Trans-disciplinary Research Grant Scheme (TRGS/1/2020/UKM/01/4/3).

Emotion regulation (ER) encompasses the activities by which humans observe, assess, adjust, and control their emotional encounters and displays. This phenomenon can manifest in both conscious and unconscious states and involves the manipulation of emotions and expression, leveraging social-emotional learning. Social-emotional learning (SEL) refers to a process that promotes the development of essential interpersonal and affective abilities in individuals [3]. SEL has gained prominence in recent years as a vital component of education and individual development.

Over the years, there has been wide research conducted on the VR app across multiple fields involving healthcare and psychology. However, there is a limited study utilizing VR that is dedicated to the development of SEL in adolescents to enhance ER. This gap is significant because adolescents face unique emotional and social challenges during their developmental stages. Effective SEL interventions during this period are crucial for fostering resilience, empathy, and interpersonal skills [4]. If VR can demonstrably boost these skills, it may result in more engaging and meaningful educational experiences that adequately prepare adolescents for real-world interactions. Therefore, we proposed a study of an app called mVR-REAL. Mobile virtual reality to read and learn emotion regulations or mVR REAL is a VR app that is designed specifically for adolescents aged between thirteen and fourteen in the public secondary school in Malaysia. The goal of developing this app was to promote social-emotional learning in adolescents to improve emotion regulation. Focusing on adolescents is crucial as this developmental stage is marked by significant emotional and social changes. During adolescence, individuals are particularly susceptible to stressors that can impact their mental health. By utilizing VR in ER and SEL practices tailored for adolescents, educators and mental health professionals can provide targeted support that addresses the unique challenges faced during this period. Engaging adolescents through immersive experiences not only enhances their learning but also promotes a deeper understanding of their emotions and social interactions.

mVR-REAL explores the vital role of SEL in fostering good mental health by using a VR technology. SEL helps promoting self-awareness, emotional regulation, interpersonal skills, and resilience, and ultimately contributing to overall mental well-being. The findings suggest the element factors of user experience on emotion regulation are emotional and psychological.

II. Literature Review

A. Virtual Reality in Healthcare

VR has been employed in diverse contexts, encompassing therapy [5], education [6], and medical [7], to help individuals regulate their emotions in real-world situations. Studies have shown that VR can be effective in reducing negative emotions [8], such as anxiety, stress, and anger, as well as increasing positive emotions, such as empathy and relaxation. VR has the capacity to completely transform and reshape various aspects of society. Emotion regulation interventions by providing a more accessible, engaging, and personalized experience for individuals. As the technology continues to advance, VR may become a more widespread and valuable tool for improving emotional well-being. It is intended to produce an experience that is so immersive and realistic that it seems very much like either the real world or a world that the user has imagined. Virtual reality can be used for a variety of purposes, including education, training, rehabilitation, and even entertainment. A key component of virtual reality is the user's ability to interact with the virtual environment [9].

As a result, virtual reality saw explosive growth across a wide range of industries. Education [10], film [11], music [12], cultural tourism [13] and medical [14] are just a few sectors where VR has already been significantly impacted. VR is an expanding area of research that has the potential to revolutionize mental health examinations [15]. VR finds applications across various industries, including gaming, education, training, healthcare, culture [13], and entertainment, offering unique opportunities for experiential learning, skill development, and creative expression [16]. With the ongoing advancement of VR technology, there is significant potential for it to bring about a revolutionary transformation in how people perceive, interact with, and experience digital content, paving the way for new and innovative forms of storytelling and human-computer interaction. VR applications have grown into a crucial component of our everyday existence, with a variety of apps available across various platforms. However, ensuring the success of a VR app in a competitive market requires meticulous testing and refinement. Preliminary testing, a vital phase in the app development process, serves as a bridge between initial development and widespread release. This study explores the significance of experimental testing for mobile VR applications, its design, methodologies, benefits, and the impact it has on user experience.

By leveraging the established benefits of VR in healthcare, educators can develop innovative SEL and ER programs that resonate with adolescents' experiences. Despite the increasing prevalence of mental health issues among youth, the incorporation of VR into educational methodologies presents a promising opportunity to improve emotional intelligence and resilience. In summary, the mVR-REAL in healthcare not only demonstrate its capacity to influence emotional responses but also provide a robust framework for its use in promoting SEL and ER among adolescents.

B. Social-Emotional Learning and Emotion Regulation

SEL enables individuals to understand and manage their emotions, build positive relationships, demonstrate empathy and compassion, and exercise prudent judgment [3]. This educational method is comprehensive and acknowledges the importance of mental agility and social consciousness in promoting both mental health and academic achievement. An organization called CASEL has established a CASEL SEL framework for developing essential life skills [17] that revolve around recognizing and managing emotions, building healthy relationships, making responsible decisions, and cultivating resilience. It is typically integrated into curricula and educational programs to promote the emotional intelligence and mental health of individuals across all age groups [18]. SEL is an essential factor in facilitating individual development, mental health, and academic achievement. By focusing on emotional intelligence and social skills [19], SEL helps students develop resilience, cope with stress, and improve problem-solving abilities [20]. It fosters a positive school climate, reduces incidents of bullying, and enhances overall classroom engagement. Furthermore, SEL is not limited to academic settings; it has broader applications in various aspects of life, including the workplace, family dynamics, and community interactions [21]. Individuals who have strong social-emotional skills are better equipped to navigate challenges, communicate effectively, and build meaningful relationships, ultimately contributing to a more compassionate and harmonious society. SEL in adolescents is a critical component of their overall development and well-being. Teenagers is a phase characterized by profound physical, emotional, and cognitive variations, and SEL helps adolescents navigate this challenging phase successfully. SEL includes a variety of abilities and qualities that empower young individuals to comprehend and handle their emotions, establish constructive connections, make accountable choices, and cultivate a robust sense of self-worth and compassion.

The study of emotion is a central area of investigation in numerous scientific and humanistic fields [22]. Emotion regulation skills are crucial for developing better well-being in the transition from adolescence to adulthood. Silvers describes emotion regulation (ER) as a set of both implicit and explicit abilities that enable individuals to observe, assess, and adjust their emotional reactions in alignment with their objectives [23]. During adolescence, the ability to regulate emotions often undergoes significant improvement for many individuals. However, for other young people, adolescence can be a time when emotion regulation difficulties begin or develop [23]. It is widely acknowledged that the capacity to control one's emotions is a crucial component of both mental [24] and general well-being [25]. A person's ability to manage their emotions, which are dynamically used throughout everyday situations, is a form of ER.

The correlation between emotion control and mental health is crucial. Emotion management is essential to maintaining good mental health and overall wellness. Deficits in the ability to adaptively deal with challenging emotions can lead to various mental disorders such as depressive disorders, borderline personality illness and eating disorders [26]. Emotion regulation tactics like rethinking thoughts and expressive suppression are related to psychological outcomes [27], such as cognitive restructuring, which involves revising a situation to alter its emotional impact, is related to better psychological health and expressive suppression, preventing the outward expression of

emotions, is linked to decreased mental wellness [27]. Adaptive strategies like suppression are associated with these adverse effects. Lack of management of emotions can lead to reduced overall well-being that includes both pleasure-oriented and meaning-oriented elements [28]. Enhancing well-being among those with mental health issues can be attained by enhancing their emotion management abilities. In short, managing emotions is a crucial mechanism that influences mental health. Focusing on emotion regulation during therapy shows promise in reducing suffering and enhancing overall quality of life [26].

C. User Experience Elements

Twenty years of emphasis on user experience (UE) has resulted in a variety of digital devices that assist individuals in perceiving, comprehending, and communicating emotions [24]. The success of a VR app depends on the quality of the content, user experience, and technical execution. The objective of the study is to investigate the UE elements of virtual reality for ER through mVR-REAL app experimental test. Preliminary testing of research protocols, also known as pilot testing, is an essential precursor to any research undertaking. This phase allows researchers to fine-tune their procedures, detect possible complications, and improve the overall integrity of their study design.

Emotional design is a concept in user experience and product design that focuses on creating products, interfaces, or environments that elicit specific emotional responses from users. The term "ER" in psychology describes a group of both conscious and unconscious methods that humans employ to diminish, sustain, or amplify positive or negative emotions [29], which leads to promoting an individual's well-being [30]. Psychological pertains to the field of psychology and behaviour, encompassing the thinking procedures and functions that impact individual and collective experiences. This field encompasses multiple facets of mental activity, emotional awareness, enthusiasm, and interactions with others, aiming to comprehend how these factors contribute to human thought patterns and actions.

Additionally, psychology can describe anything related to the mental and emotional aspects of an individual or group, including the effects of thoughts and feelings on behaviour and well-being [31]. In the context of user experience (UE) design, psychological elements refer to aspects of human cognition, emotion, and behaviour that influence how individuals perceive and interact with a product, service, or interface. Understanding these psychological elements is crucial for creating designs that are user-friendly, engaging, and effective. Psychological principles perform an essential function in UE design by impacting how users perceive, interact with, and respond to digital products and interfaces. The element factors of user experience on emotion regulation, specifically focusing on the emotional and psychological aspects, suggest significant correlations between the UE of the app and its impact on users' emotional and psychological well-being [32]. The emotional design of a user interface, including visual elements and colour schemes plays a crucial role in influencing users' emotional states [33].

Cultural aspects in app design are critical considerations that go beyond mere aesthetics [34]. They encompass a deeper understanding and integration of cultural elements to create an inclusive and user-friendly experience for individuals from diverse backgrounds. For instance, in Malaysia uses the Malay language as the first language. In the consideration of building the app, we have selected the Malay language as the language used in the whole app representation. We can create products that resonate with a global audience, fostering positive user experiences across diverse cultural contexts through integrating these cultural considerations into the app design process.

III. METHODOLOGY

A. Participants

Participating in the study are sixteen students who are currently enrolled in secondary school in a public school in Malaysia. The subject's range in age from thirteen to fourteen years old during the study.

B. Ethics Statement

The study was conducted on March 17, 2023, comprising eight males and eight females, all fourteen-year-old secondary school students. Informed consent was obtained from all participants, and because the study involved minors, written parental consent was also secured from their parents or guardians. Participants were monitored for emotional well-being throughout the intervention, with arrangements in place to refer any identified emotional issues to the school counsellor. All data collected during the study were securely stored and maintained under strict confidentiality for the entire duration of the research. This study has earned over the Research Ethics Committee's approval, The National University of Malaysia with the ethics approval number, UKM PPI/111/8/JEP-2023-111. The Malaysian Ministry of Education has also given its permission (KPM.600-3/2/3-eras (15996)).

C. Instruments

The study involved a VR tool named Meta Quest 2 (MQ2) previously known as Oculus Quest 2. MQ2 is a virtual reality headset developed by Oculus, which is a subsidiary of Facebook which now Meta Platforms, Inc. It's a standalone VR headset that offers wireless and untethered virtual reality experiences. The MQ2 features improved hardware specifications, higher resolution displays, and enhanced processing power compared to its predecessor, the original Oculus Quest. There were three Meta Quest 2 headsets that is used by one per participant. We first pre-installed mVR-REAL apps for the pilot test. We conducted a pilot/feasibility study as an initial measure towards subsequently carrying out a comprehensive trial [35].

Participants are generally given explicit instructions on how to interact with the VR experience. There is also an initial tutorial that guide participant on how to navigate the VR space and use controllers effectively. The participants were given a briefing regarding the apps which includes a questionnaire administered before and after a test. Comparison of pre- and post-test surveys are valuable tools for assessing changes in knowledge, attitudes, behaviours, or other variables before and after an intervention.

The pre-test comprising fifteen questions on the effectiveness of the apps, and the same set of fifteen questions were handed out as a post-test questionnaire to evaluate the app's

efficacy once again after the VR session. The initial step involved conducting a systematic literature review (SLR) on VR for SEL to identify key themes, gaps, and findings in existing research related to VR applications in SEL and ER. From the SLR, specific elements that consistently emerged as crucial for effective SEL and ER interventions were identified. These included emotional awareness, coping strategies, social interactions, and empathy development. The questions were derived from a prior SLR to investigate these elements further, ensuring relevance to the study's objectives. The questionnaire conducted based on Alqahtani's method represents a valuable approach to gaining insights into various aspects of user experience [36]. Participants were given the opportunity to use the mVR-REAL after completing a pre-test questionnaire. The session took place in the school auditorium. A VR element that lasts approximately three to five minutes per scene is included in the episode. Pre and post-test questionnaires are valuable tools for assessing changes in knowledge, attitudes, behaviours, or other variables before and after an intervention.

Table I indicates a set of question for pre and post-test questionnaire. The questions are designed based on an element that found in the previous study of the systematic literature review [37]. The answer provides a multiple-choice Likert scale namely, strongly agree, agree, neutral, disagree and strongly disagree. A value in numbers was allocated to each notional choice on the scale. A Likert scale is a prevalent rating system used to measure attitudes, opinions, or perceptions [38]. It typically consists of a statement followed by a range of response options that reflect varying degrees of agreement or disagreement.

TABLE I. PRE AND POST-TEST QUESTIONNAIRE

Number	Question
1.	Sound plays a role in understanding something that is not clear
2.	Pictures make a design interesting
3.	Certain methods are required to convey information
4.	Effective interaction affects the presentation of a good story
5.	The physiology of body movement helps me act better
6.	I believe facial expressions and gestures are the result of understanding something
7.	I do not really know how to deal with stress
8.	I always hide my fear
9.	I often cannot control my anger
10.	I believe I have a wide variety of emotions
11.	I believe my behavior is influenced by my emotions
12.	I will show happiness if I feel calm
13.	I believe peace can make me happy

D. Data Analysis

The SPSS, short for Statistical Package for the Social Sciences, is a highly popular software application utilized extensively in the realm of research and data analysis. SPSS offers an interface that is easy to use and facilitates data input, manipulation, analysis, and visualization, all of which contribute to the successful discovery of insights. The acquired data was analysed using the SPSS method and paired sample t-test. The paired sample t-test was selected for this investigation because it is appropriate for comparing the means of two related groups, specifically measurements obtained from the same subjects before and after an intervention. Comparison of pre and post-test surveys provide a robust method for evaluating the impact of interventions and tracking changes in variables over time. The t-test result revealed statistically significant differences in the results users experienced after interacting with the mVR-REAL The apps enhance the user experience and facilitate participants' comprehension of emotion regulations.

IV. FINDINGS

A. User Experience Impact Based on Paired Sample T-Test

The paired samples t-test was conducted to evaluate the effectiveness of the VR intervention on SEL and ER among adolescents. The key results include the means, standard deviations, and t-values, which provide insights into the impact of the intervention. By evaluating the average scores from the pre and post-tests, a paired samples t-test was performed.

Table II shows the outcomes of the descriptive analysis using matched sample statistics on the data. The median number represents the calculated mean of the data. The desired amount might be derived over summing all the numbers provided in the data and thereafter dividing the result by the total number of variables.

TABLE II. MATCHED SAMPLE DATA

Matched Samples Data					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Pre-test	3.7417	15	.60774	.15692
	Post-test	3.9625	15	.44809	.11570

The mean score represents the average performance of participants on SEL and ER assessments before and after the VR intervention. The mean score in the pre-test is 3.74 and median score during the post-test is 3.96, this indicates an improvement in participants' skills.

The standard deviation measures how much individual scores deviate from the mean. A smaller standard deviation suggests that participants' scores are closely clustered around the mean, while a larger standard deviation indicates more variability in responses. For instance, the standard deviation before the intervention was .60774 and after was .44809, which suggests that participants became more consistent in their scores following the VR experience. The following demonstrates the fact that there are distinct variations in the descriptive results obtained from user experiences.

The variable N represents the quantity of information collected for every parameter. In the current study, a maximum of fifteen participants were included in the research sample, denoted by the value of N in the table. The variation of the data around the median is revealed by the typical variation. Data is considered more diverse when its standard deviation is large, and inversely. The mean is not a good indicator of the data quality if the standard variance is higher than the mean. The median value is an accurate reflection of the complete data set if

and only if the average variation value is lower than the mean value. Eq. (1) can be used to determine the average variation explicitly.

$$s = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}} \quad (1)$$

The symbol S represents the usual variation, $\bar{X}$ represents the median amount of the data, and n represents the number of data points. The findings indicate that the median variation among the pre-test is 0.60774 and the post-test is 0.44809. This indicates that both the variances are lower than the mean value. Hence, the mean of the pre-test and post-test may accurately portray the entire dataset. Eq. (2) provides a method to calculate the mean standard error ($s\bar{x}$) through dividing the standard deviation by the square root of the population length.

$$S\bar{x} = \frac{s}{\sqrt{n}} \quad (2)$$

The variable s represents the standard deviation, whereas the variable n represents the number of data points. The mean standard error provides a measure of how accurately the mean data across each sample can approximate the overall means for every factor.

TABLE III. PAIRED T-TEST RESULTS

Paired Sample Test								
Pair 1	Paired Changes					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Break Difference				
				Lower	Upper			
Pre-test - Post-test	-.22083	.24306	.06276	-.35543	-.08623	-3.519	14	.003

Table III displays the outcomes of a paired sample t-test. The median figure of Table III is the mean of the bilateral variances generated by the variance between the median values. Test outcomes indicate that the median measurement for the variance is -.22083. The stated average variance is the total standard variation calculated from the differences between pairs. The average variation of the variance can be computed using the previously given equation [39].

The test findings indicate an average of the difference of 0.24306. The median average error is a statistical measure that provides an estimation of the normal variation over the representative standard probability for a population that is infinitely large. It is possible to determine the median error of the average variation using the formula provided. We know that the median score has a standard error of.06276. The range of certainty for the mean difference has upper and lower bounds that represent the confidence interval's confidence.

In summary, the results from the paired samples t-test reveal that participants showed significant improvements in their SEL and ER skills after engaging with the VR intervention. The mean scores increased, indicating better performance, while changes in standard deviation suggest greater consistency among participants' responses post-intervention. The statistical significance ($p < 0.05$) confirms that these improvements are unlikely to be due to chance, and a moderate effect size indicates meaningful benefits from using VR for enhancing emotional and social competencies among adolescents. This interpretation provides a clear understanding of how statistical measures reflect the effectiveness of the VR intervention without unnecessary complexity or redundancy.

This demonstrates the utilization of mVR-REAL apps has a significant influence on user experience outcomes. mVR-REAL apps are transforming the way users engage with digital content. Through an exploration of user immersion, satisfaction, and learning, this study reveals the profound impact of mVR-REAL apps on enhancing user engagement and learning in a virtual environment. The findings suggest the element factor of UE on ER are emotional and psychological. The construct will guide with the matrix on testing the mVR-REAL app.

TABLE IV. DESCRPTIVE ANALYSIS

Tests	*n*	Descriptive Statistics	Paired T-Test		
		M (Std. D)	*t*	*df*	Sig. (2-tailed)
Pre-test	15	3.74 (0.60)	-3.519	14	0.003*
Post-test	15	3.96 (0.44)			

Table IV presents a descriptive statistic. Aside from the t-value, relevance can also be inferred from the p-value or significant value seen in Table IV. The decision-making process is based on the criterion that if the significance (Sig. (2-tailed)) is below 0.05, the null hypothesis (H0) which states that there is no relationship between two paired groups, is rejected, and the alternative hypothesis (Ha) which states that there is a relationship between the two paired groups, is accepted. The paired samples t-test yielded a significance level (Sig. (2-tailed)) of 0.003, leading to the rejection of the null hypothesis (H0) and acceptance of the alternative hypothesis (Ha). Therefore, the utilization of mVR-REAL apps has a significant influence on UE outcomes which signifying an element of emotion and psychological. This displays that mVR-REAL apps are effectively used in regulating emotions among participants. The findings suggest the element factor of UE on ER are emotions and psychological. The construct will guide with the matrix on testing the mVR-REAL app.

V. DISCUSSION

The findings of this study underscore the significant potential of VR as an innovative tool for enhancing ER practices, aligning with existing literature that emphasizes the importance of immersive environments in promoting emotional well-being. Previous research has shown that VR can create a heightened sense of presence and engagement, which are critical for effective emotional interventions. This study builds on that foundation by demonstrating how specific design elements such

as aesthetics, audio, and personalization can further enhance the effectiveness of VR in supporting ER.

A. User Experience Element of Virtual Reality for Emotion Regulation

According to the findings, mVR-REAL apps positively impact the user experience. The findings propose two important UE element needed in developing a VR app to enhance ER namely emotional and psychological. In the development of a VR app aimed at enhancing emotion regulation, two crucial user experience elements are the consideration of emotional design and the incorporation of psychological principles. Both elements play significant roles in creating an immersive and effective experience for users seeking to regulate their emotions.

1) Emotional design

a) Aesthetic and visual elements: The visual aspects of a VR environment have a profound impact on users' emotions. The choice of colours, textures, and overall aesthetics can evoke specific emotional responses. For emotion regulation, calming and soothing visual elements, such as nature scenes or soft colour palettes, may be incorporated to create a positive and serene atmosphere.

b) Immersive audio: Sound and music are powerful tools for influencing emotions. Integrating immersive audio experiences that align with the desired emotional state can enhance the overall effectiveness of the app. This could involve calming ambient sounds, guided meditations, or music designed to induce relaxation.

c) Personalization: Allowing users to personalize their virtual environment can contribute to a sense of control and emotional comfort. Customizable elements, such as avatar appearance or the choice of surroundings, empower users to tailor the experience to their preferences, contributing to a more emotionally resonant interaction.

2) Psychological principles

a) Cognitive behavioural techniques: Incorporating principles from cognitive-behavioural therapy (CBT) into the VR app can provide users with practical tools for managing their emotions. This may involve interactive exercises, guided cognitive restructuring, and behaviour modification strategies that users can engage with in the VR environment. This information can be used to guide users through exercises aimed at regulating their emotional and physiological states, promoting self-awareness and control. The integration of CBT and mindfulness within VR interventions is particularly compelling due to the unique enhancements provided by the immersive nature of VR. Compared to traditional methods, VR offers several advantages that significantly bolster the effectiveness of these therapeutic techniques.

*b) Social skills***:** Implementing features that allow users to learn about their social skill to enhance the psychological impact of the app. Progress tracking, along with reflective exercises, can contribute to a sense of accomplishment and self-discovery, reinforcing positive emotional experiences.

*c) Guided mindfulness***:** Integrating mindfulness practices and relaxation techniques into the VR app such as an exercise can help users cultivate a focused and present state of mind. Guided sessions led by virtual mentors or interactive experiences that encourage mindfulness can contribute to emotional regulation by promoting a sense of calm and centeredness.

3) Cultural aspect

a) Content representation: Consideration should be given to the representation of diverse cultures of Malaysia within the mVR-REAL content. This includes incorporating scenarios, environments, and characters that are inclusive and representative of Asian cultural contexts. Users from different cultural backgrounds should feel a sense of inclusivity and relevance in the virtual experiences.

b) Language and communication: The app's language and communication should be mindful of cultural nuances. mVR-REAL involves using Malay language and ensuring that any written or spoken content is easily understandable and culturally appropriate for a Malaysian user base. Language choices should be respectful and considerate of cultural diversity. Therefore, mVR REAL chose Malay language in the whole app presentation because Malay language is close to heart of Malaysian. Emphasizing cultural representation and language in the context of VR applications is crucial for enhancing user engagement and ensuring that emotional and psychological benefits are universally accessible, particularly within the Malaysian context. Cultural relevance not only enriches the user experience but also fosters a deeper connection between the content and the users, leading to more meaningful interactions.

In summary, the findings of this study highlight the critical importance of integrating emotional, psychological, and cultural elements in the development VR applications for ER. By focusing on these aspects, developers can create immersive experiences that resonate deeply with users, enhancing engagement and therapeutic outcomes. Ultimately, mVR-REAL applications hold promise as effective tools for persons aiming to improve their emotional well-being and regulation by immersing themselves in engaging virtual environments. The findings suggest that the mVR-REAL app effectively integrates emotional, psychological, and cultural elements to enhance emotion regulation. Users experienced positive emotional effects, benefitted from cognitive-behavioural techniques, and appreciated the cultural sensitivity and inclusivity of the app, contributing to an overall positive user experience in the pursuit of improved emotional well-being. The mVR-REAL app that incorporates various elements of emotional design, psychological principles and cultural aspect has successfully impacted the user experience. This investigation has illuminated on the positive impact of various element features on user experience across a wide range of digital and physical environments involving VR apps. Furthermore, the findings from this study provide valuable insights for VR developers aiming to create applications that effectively support ER. By prioritizing emotional design and psychological principles during the early stages of development, developers can enhance user engagement and improve therapeutic outcomes.

Further research will be directed towards the concrete implementation of usability testing for the mVR-REAL project

that should focus on several key questions that emerge from the findings of this study such as how do different design elements affect user satisfaction and engagement in VR environments for ER? and what is the impact of culturally tailored content on the effectiveness of VR interventions for diverse populations?. Moreover, challenges related to time were also identified, necessitating further research on how does the duration of VR sessions influence the effectiveness of ER techniques?

In a nutshell, as the technology continues to evolve, it is imperative for developers and researchers to explore innovative ways to harness the potential of mVR-REAL apps, ultimately providing users with immersive and transformative digital experiences. The app holds the promise of reshaping how we learn, entertain, and engage with digital content in the mobile VR era. The potential long-term impact of mVR-REAL apps on the field of ER is substantial, with implications that extend beyond VR app development into broader digital health interventions. By leveraging immersive technology, these applications can transform how individuals manage their emotions, offering innovative solutions that are both engaging and effective. mVR-REAL apps can cultivate a deeper understanding of emotional states through immersive experiences that allow users to visualize and interact with their emotions in real-time. This heightened awareness can lead to improved emotional regulation skills over time, fostering resilience and adaptability in various life situations.

ACKNOWLEDGMENT

I. Hamzah expresses gratitude to the Ministry of Education, Malaysia for their crucial assistance in funding this work, provided under the Trans-disciplinary Research Grant Scheme (TRGS/1/2020/UKM/01/4/3).

REFERENCES

[1] I. Wohlgenannt, A. Simons, and S. Stieglitz, "Virtual Reality," Business & Information Systems Engineering, 2020, doi: 10.1007/s12599-020-00658-9.

[2] C. Girvan, "What is a virtual world? Definition and classification," Educational Technology Research and Development, vol. 66, no. 5, pp. 1087–1100, 2018, doi: 10.1007/s11423-018-9577-y.

[3] C. N. Ly and M. Soutter, "The Collaborative Model for Teaching o-SEL: Preparing Educators to Design Online Environments for Social-Emotional Learning," vol. 8, pp. 71–87, 2022, [Online]. Available: https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-7998-9168-0.ch004.

[4] A. A. Colomeischi, A. Ursu, I. Bogdan, A. Ionescu-Corbu, R. Bondor, and E. Conte, "Social and Emotional Learning and Internalizing Problems among Adolescents: The Mediating Role of Resilience," Children, vol. 9, no. 9, Sep. 2022, doi: 10.3390/children9091326.

[5] B. M. Maarsingh, J. Bos, C. F. J. Van Tuijn, and S. B. Renard, "Changing Stress Mindset Through Stressjam: A Virtual Reality Game Using Biofeedback," Games Health J, vol. 8, no. 5, pp. 326–331, 2019, doi: 10.1089/g4h.2018.0145.

[6] R. Liu, L. Wang, J. Lei, Q. Wang, and Y. Ren, "Effects of an immersive virtual reality-based classroom on students' learning performance in science lessons," British Journal of Educational Technology, vol. 51, no. 6, pp. 2034–2049, 2020, doi: 10.1111/bjet.13028.

[7] N. Elmqaddem, "Augmented Reality and Virtual Reality in education. Myth or reality?," International Journal of Emerging Technologies in Learning, vol. 14, no. 3, pp. 234–242, 2019, doi: 10.3991/ijet.v14i03.9289.

[8] P. Lindner, W. Hamilton, A. Miloff, and P. Carlbring, "How to treat depression with low-intensity virtual reality interventions: Perspectives on translating cognitive behavioral techniques into the virtual reality modality and how to make anti-depressive use of virtual reality–unique experiences," Front Psychiatry, vol. 10, no. OCT, pp. 1–6, 2019, doi: 10.3389/fpsyt.2019.00792.

[9] J. Torous et al., "The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality," World Psychiatry, vol. 20, no. 3, pp. 318–335, 2021, doi: 10.1002/wps.20883.

[10] E. McGovern, G. Moreira, and C. Luna-Nevarez, "An application of virtual reality in education: Can this technology enhance the quality of students' learning experience?," Journal of Education for Business, vol. 95, no. 7, pp. 490–496, Oct. 2020, doi: 10.1080/08832323.2019.1703096.

[11] J. Mateer, "Directing for Cinematic Virtual Reality: how the traditional film director's craft applies to immersive environments and notions of presence," Journal of Media Practice, vol. 18, no. 1, pp. 14–25, Jan. 2017, doi: 10.1080/14682753.2017.1305838.

[12] X. Lin et al., "Virtual Reality-Based Musical Therapy for Mental Health Management," in 2020 10th Annual Computing and Communication Workshop and Conference, CCWC 2020, Institute of Electrical and Electronics Engineers Inc., Jan. 2020, pp. 948–952. doi: 10.1109/CCWC47524.2020.9031157.

[13] A. Marasco, "Beyond virtual cultural tourism: history-living experiences with cinematic virtual reality," Tourism and Heritage Journal, vol. 2, pp. 1–16, May 2020, doi: 10.1344/thj.2020.2.1.

[14] S. H. M. Chan, L. Qiu, G. Esposito, K. P. Mai, K. P. Tam, and J. Cui, "Nature in virtual reality improves mood and reduces stress: evidence from young adults and senior citizens," Virtual Real, 2021, doi: 10.1007/s10055-021-00604-4.

[15] I. H. Bell, J. Nicholas, M. Alvarez-Jimenez, A. Thompson, and L. Valmaggia, "Virtual reality as a clinical tool in mental health research and practice," Dialogues Clin Neurosci, vol. 22, no. 2, pp. 169–177, 2020, doi: 10.31887/DCNS.2020.22.2/lvalmaggia.

[16] V. Hohmann, R. Paluch, M. Krueger, M. Meis, and G. Grimm, "The Virtual Reality Lab: Realization and Application of Virtual Sound Environments," Ear Hear, vol. 41, pp. 31S-38S, 2020, doi: 10.1097/AUD.0000000000000945.

[17] C. Dermody et al., "A Developmental Framework for the Integration of Social and Emotional Learning and Career and Workforce Development," Casel, pp. 1–10, 2022, [Online]. Available: https://casel.org/sel-workforce-brief-03-2022/.

[18] D. Donahue-Keegan, E. Villegas-Reimers, and J. M. Cressey, "Integrating Social-Emotional Learning and Culturally Responsive Teaching Integrating Social-Emotional Learning and Culturally Responsive Teaching in Teacher Education Preparation Programs: The Massachusetts Experience So Far," 2019.

[19] K. M. Murphy, A. L. Cook, and L. M. Fallon, "Mixed reality simulations for social-emotional learning," Phi Delta Kappan, vol. 102, no. 6, pp. 30–37, 2021, doi: 10.1177/0031721721998152.

[20] D. Katz, J. Mahfouz, and S. Romas, "Creating a Foundation of Well-being for Teachers and Students Starts with SEL Curriculum in Teacher Education Programs," Northwest Journal of Teacher Education, vol. 15, no. 2, Oct. 2020, doi: 10.15760/nwjte.2020.15.2.5.

[21] K. A. Schonert-Reichl, "Advancements in the Landscape of Social and Emotional Learning and Emerging Topics on the Horizon," 2019, Routledge. doi: 10.1080/00461520.2019.1633925.

[22] G. Wadley et al., "The Future of Emotion in Human-Computer Interaction," in Conference on Human Factors in Computing Systems - Proceedings, Association for Computing Machinery, Apr. 2022. doi: 10.1145/3491101.3503729.

[23] J. A. Silvers, "Adolescence as a pivotal period for emotion regulation development for consideration at current opinion in psychology," Apr. 01, 2022, Elsevier B.V. doi: 10.1016/j.copsyc.2021.09.023.

[24] W. Smith et al., "Digital Emotion Regulation in Everyday Life," in Conference on Human Factors in Computing Systems - Proceedings, Association for Computing Machinery, Apr. 2022. doi: 10.1145/3491102.3517573.

[25] Y. Shi, P. Koval, V. Kostakos, J. Goncalves, and G. Wadley, "'Instant Happiness': Smartphones as tools for everyday emotion regulation,"

International Journal of Human Computer Studies, vol. 170, Feb. 2023, doi: 10.1016/j.ijhcs.2022.102958.

[26] M. Berking and P. Wupperman, "Emotion regulation and mental health: Recent findings, current challenges, and future directions," Mar. 2012. doi: 10.1097/YCO.0b013e3283503669.

[27] T. Hu, D. Zhang, J. Wang, R. Mistry, G. Ran, and X. Wang, "Relation between emotion regulation and mental health: A meta-analysis review," Psychol Rep, vol. 114, no. 2, pp. 341–362, 2014, doi: 10.2466/03.20.PR0.114k22w4.

[28] J. T. Kraiss, P. M. ten Klooster, J. T. Moskowitz, and E. T. Bohlmeijer, "The relationship between emotion regulation and well-being in patients with mental disorders: A meta-analysis," Compr Psychiatry, vol. 102, Oct. 2020, doi: 10.1016/j.comppsych.2020.152189.

[29] S. Balzarotti, F. Biassoni, D. Villani, A. Prunas, and P. Velotti, "Individual Differences in Cognitive Emotion Regulation: Implications for Subjective and Psychological Well-Being," J Happiness Stud, vol. 17, no. 1, pp. 125–143, Feb. 2016, doi: 10.1007/s10902-014-9587-3.

[30] J. C. Stevenson, A. Millings, and L. M. Emerson, "Psychological Well-being and Coping: the Predictive Value of Adult Attachment, Dispositional Mindfulness, and Emotion Regulation," Mindfulness (N Y), vol. 10, no. 2, pp. 256–271, Feb. 2019, doi: 10.1007/s12671-018-0970-8.

[31] J. T. Kraiss, P. M. ten Klooster, J. T. Moskowitz, and E. T. Bohlmeijer, "The relationship between emotion regulation and well-being in patients with mental disorders: A meta-analysis," Compr Psychiatry, vol. 102, Oct. 2020, doi: 10.1016/j.comppsych.2020.152189.

[32] T. Partala and T. Saari, "Understanding the most influential user experiences in successful and unsuccessful technology adoptions," Comput Human Behav, vol. 53, pp. 381–395, Jul. 2015, doi: 10.1016/j.chb.2015.07.012.

[33] A. Gaggioli, G. Riva, D. Peters, and R. A. Calvo, "Positive Technology, Computing, and Design: Shaping a Future in Which Technology Promotes Psychological Well-Being," in Emotions and Affect in Human Factors and Human-Computer Interaction, Elsevier, 2017, pp. 477–502. doi: 10.1016/B978-0-12-801851-4.00018-5.

[34] F. Zheng and S. Hermawati, "Cultural Aspects for the User Interface Design of Health and Fitness Apps," 2017.

[35] G. A. Lancaster, "Pilot and feasibility studies come of age!," Apr. 01, 2015, BioMed Central Ltd. doi: 10.1186/2055-5784-1-1.

[36] A. Alqahtani, "Usability testing of google cloud applications: Students' perspective," J Technol Sci Educ, vol. 9, no. 3, pp. 326–339, 2019, doi: 10.3926/JOTSE.585.

[37] I. Hamzah et al., "Virtual Reality for Social-Emotional Learning: A Review," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Science and Business Media Deutschland GmbH, 2024, pp. 119–130. doi: 10.1007/978-981-99-7339-2_11.

[38] J. Mumu, B. Tanujaya, R. Charitas, and I. Prahmana, "Likert Scale in Social Sciences Research: Problems and Difficulties," FWU Journal of Social Sciences, vol. 16, no. 4, pp. 89–101, 2022, doi: 10.51709/19951272/Winter2022/7.

[39] S. Afifah, A. Mudzakir, A. Bayu, and D. Nandiyanto, "How to Calculate Paired Sample t-Test using SPSS Software: From Step-by-Step Processing for Users to the Practical Examples in the Analysis of the Effect of Application Anti-Fire Bamboo Teaching Materials on Student Learning Outcomes Indonesian Journal of Teaching in Science," Indonesian Journal of Teaching in Science, vol. 2, no. 1, pp. 81–92, 2022, doi: 10.17509/xxxx.xxxx.

Enhancing Emotion Regulation Through Virtual Reality Design Framework for Social-Emotional Learning (VRSEL)

Irna Hamzah[1], Ely Salwana[2], Nilufar Baghaei[3], Mark Billinghurst[4], Azhar Arsad[5]
Institute of Visual Informatics, Universiti Kebangsaan Malaysia (UKM), Bangi, Selangor, Malaysia[1, 2]
School of Electrical Engineering and Computer Science, The University of Queensland, St Lucia, Australia[3]
Auckland Bioengineering Institute, The University of Auckland, Auckland 1010, New Zealand[4]
V3X Malaysia Sdn. Bhd., 87A, Jalan Pulai 7, Taman Pulai Utama, 81300 Skudai, Johor, Malaysia[5]

***Abstract*—Virtual reality (VR) has swiftly progressed, transitioning from a niche technology primarily associated with gaming to a versatile tool with broad applications across entertainment, healthcare, education, and beyond. Social-emotional learning (SEL) is increasingly recognized for its role in enhancing individuals' social skills and emotional regulation. However, despite the growing body of research on VR, the development of specific design frameworks for integrating VR with SEL remains underexplored. This study addresses this gap by employing thematic analysis to identify the critical components necessary for a Virtual Reality Design Framework for Social-Emotional Learning (VRSEL) aimed at improving emotion regulation among Malaysian adolescents. Through qualitative data derived from expert interviews in SEL and VR, this research proposes a framework that leverages immersive VR technology to create realistic, interactive scenarios that facilitate the practice and development of social-emotional skills. The framework emphasizes key design principles, including user interface (UI), presentation layer (PL), and brain activity (BA). Our findings suggest that VRSEL is a powerful tool for SEL, offering significant potential for educational environments. However, challenges such as technical barriers, content development, and educator training must be addressed to fully realize its benefits. This research highlights the promising role of VR in advancing SEL and lays the groundwork for further exploration and refinement of VRSEL in diverse educational settings.**

Keywords*—*Virtual reality; design; framework; social-emotional learning; emotion regulation

I. Introduction

Virtual reality (VR) can be described as a computer-generated simulation technology which allows a user to interact with a computer-simulated environment [1] by employing specialized electronic apparatus [2]. Modern virtual reality (VR) technology commonly utilizes head-mounted displays (HMDs), also known as VR headsets, which allow users to immerse themselves in a virtual environment by obscuring their perception of the actual world [3]. VR offers a novel setting in users to experience being immersed. Immersion is the state where a user becomes deeply engaged in a virtual environment, leading to a detachment from the reality of time and their physical surroundings [4]. The immersive experience enables audiences to completely diverge from the actual world. Through the provision of well-defined concepts, elements, and methods, it ensures consistency, fosters collaboration, and enables the growth of design efforts. VR represents the integration of hardware and software systems to generate the accurate and immersive experience of being physically present in another dimension. VR design entails the development of immersive digital environments that can effectively and realistically engage with users. To achieve this, it is crucial to possess a comprehensive design framework.

Partelow emphasizes that, as stated by The Cambridge Dictionary, frameworks are "a structural foundation upon which something can be constructed; a set of regulations, concepts, or principles used for organizing or determining something" [5]. A design framework is a structured approach used to direct the design process throughout several stages, including software and app development and more. A design framework is a concise and visually structured system that enhances the organization of information and ideas related to an issue, resulting in improved effectiveness. Frameworks are visual or written representations of the elements, ideas, or variables of a phenomenon [6]. This framework facilitates the development process, guaranteeing that the result is captivating, operational, and easily reachable. The process of designing and developing software or any other product can be intricate. In the presence of several obstacles, we are required to have a have a well-defined framework and methodologies to effectively conceptualize, construct, evaluate, and refine optimal ideas. Through the adoption of frameworks, we can enhance their comprehension of problems and discern alternate strategies and answers that may not be immediately evident. Frameworks provide a systematic approach to organizing the process, igniting creativity, and promoting cooperation. They are fundamental components that direct the development of design solutions. Frameworks help structure the process, spark innovation, and stimulate collaboration which are essential building blocks for guiding design solutions. Several VR design frameworks are utilized in diverse fields, including healthcare [7] and education [8].

Social-emotional learning (SEL) is defined as the process through which individuals, particularly students, learn to recognize and control their emotions [9], develop empathy for others, establish positive relationships, and make responsible decisions [10]. It encompasses a set of skills that are essential

The study was funded by Malaysia's Ministry of Education under Trans-disciplinary Research Grant Scheme (TRGS/1/2020/UKM/01/4/3).

for personal and social development, including emotional awareness, emotional regulation, social awareness, relationship skills and decision making. SEL is crucial for fostering not only emotional intelligence but also academic success. Programs that effectively implement SEL strategies have been shown to improve students' social skills, enhance their academic performance, and reduce behavioural problems [11]. As SEL widens its attention to include teenage populations, it adopts the widely acknowledged theoretical framework proposed by an organization called CASEL or the Collaborative for Academic, Social, and Emotional Learning [12]. CASEL has released a framework to help spread SEL curriculum widely [13]. This framework organizes SEL competencies and helps identify well-designed, evidence-based SEL programs in a systematic manner. Schools strive to foster a nurturing atmosphere that enhances students' emotional well-being and academic performance by incorporating SEL into their educational framework. The ultimate objective is to provide students with the essential abilities and mindset to achieve outstanding academic performance and prosper in several other areas of life [13]. SEL is an essential element of education involves cultivating abilities such as empathy and emotional regulation.

Emotion regulation (ER) refers to the processes of individuals manage or control their emotional experiences and expression [14], modifying the initiation or length of emotions [15], and emotional modulation [16]. The area of emotion regulation is experiencing rapid growth within the realm of psychology [17]. Through the assistance of ER, individuals can manage their emotions, when they feel it, and how they express them. This includes both conscious and unconscious techniques. Being able to regulate one's emotions is crucial for thriving in social settings when letting one's guard down is generally frowned upon. ER is a regular and necessary process for maintaining wellbeing, and it is used flexibly in various everyday situations [18].

VR has been a rapidly developing technology with various potential applications, including in the field of psychology and mental health to treat depression and anxiety [19]. However, there is limited research that focus on enhancing emotion regulation in adolescent using VR. Therefore, the goal of this study is to develop a virtual reality design framework for social-emotional learning (VRSEL) that enhances emotion regulation in Malaysian adolescents. VR is being actively promoted as a highly promising technique for teaching and learning in various educational settings [20]. The utilization of VR offers an additional chance to effectively improve SEL through immersive educational experiences. This is accomplished by promoting improved regulation of emotions, which contributes to an individual's general state of mental health. Thematic analysis is employed to utilize qualitative data from interviews with experts in the creation of VRSEL. Thematic analysis (TA) is a systematic approach used to find, analyse, and interpret patterns of meaning or themes within qualitative data [21]. It entails a methodical procedure of analysing and reviewing data, such as transcripts of interviews or discussions in focus groups, to identify important themes that encapsulate the core of the dataset. The VRSEL framework will provide a vital guideline for other researchers to develop a specific framework in the context of emotion regulation by promoting SEL.

The aim of this study is to design a specific framework that integrate VR with SEL to enhance emotion regulation in Malaysian adolescents. This study fills this gap by utilizing thematic analysis to discern the essential elements required for VRSEL framework, applying qualitative data obtained from expert interviews in SEL and VR.

II. Literature Review

A. Virtual Reality Design Framework

Design frameworks are essential for directing the developing phase and enabling successful creation of goods or systems that prioritize the needs of the user. The application of design frameworks has been explored in various domains, including user experience design, service design, and interaction design, demonstrating their versatility and effectiveness in addressing different design challenges. The field of VR design has evolved significantly in recent years that focused on refining VR design frameworks to improve usability [22], user experience [23], and the overall effectiveness [24] of VR apps across various domains.

1) Improvement of design skills: The potential of virtual reality for strengthening design skills is rapidly being acknowledged. Studies suggest that immersive settings promote embodied experiences, leading to improved creative problem-solving and design-thinking abilities [25].

2) Consumer assessment in VR: Consumer examination of items in VR environments is another important field of study. A review has highlighted key characteristics related to how customers evaluate items in VR, indicating that VR can be a valuable platform for marketers to conduct testing and prototype development (Branca et al., 2024).

3) Skill training frameworks: An extensive analysis of VR characteristics for skill training emphasized the importance of using VR technology in specific contexts (Tusher et al., 2024). The study highlighted that the efficacy of VR in training situations relies on the congruence between VR attributes and desired learning objectives.

4) Unified adoption framework: A comprehensive investigation, including both qualitative and quantitative methodologies, has put forward a novel paradigm for the unified implementation of VR (Fares et al., 2024). This review investigates the factors that impact consumer involvement with VR technology. The primary objective of this framework is to offer a thorough comprehension of the elements that either help or impede the adoption of VR in the context of design.

In the education sector, VR frameworks aim to enhance learning experiences through immersive environments that promote engagement and retention. Educational VR applications often incorporate gamification elements to make learning more interactive and enjoyable (Lyu et al., 2023). The integration of VR in education has shown promising results in various domains, including science, history, and language learning (McGovern et al., 2020). However, its application in SEL is relatively underexplored. Prior research suggests that immersive VR experiences can significantly impact users' emotional and social skills by placing them in realistic,

emotionally charged scenarios (Marín-Morales et al., 2020). This study builds on this foundation to create a targeted framework for SEL.

B. Human Computer Interaction and Design Principle

Human-Computer Interaction (HCI) is a discipline that focuses on the optimization and efficacy of user interfaces in relation to computer systems, user-friendly input and output technology, and the psychological elements of user interfaces [26]. The domain of HCI has broadened to encompass a diverse array of devices and technologies, such as mobile devices, the Internet of Things, and virtual reality. VR technology, along with HCI, plays a crucial role in software development by minimizing errors and achieving higher levels of accuracy and quality.

This is achieved through the utilization of design principles that specifically target the user interface and presentation layer. User Interface (UI) design is a critical aspect of HCI that focuses on the look and feel, responsiveness, and interactivity of digital products. Efficient UI design guarantees that people can interact with systems effectively and provides a visually appealing presentation layer. The presentation layer (PL) of HCI refers to how information is displayed to users. UI is greatly influenced by several key components, including information architecture, visual hierarchy, and responsive design. The design of the presentation layer should prioritize the user's needs, ensuring that it is both aesthetically pleasing and practical, with a user-friendly interface. Research highlights the need of adaptive interfaces that can adjust to user context and preferences, hence improving usability and pleasure [27].

C. Social-Emotional Learning and Emotion Regulation

Twenty SEL and emotion regulation are foundational to students' academic success, school readiness, and overall well-being [28], [29]. SEL involves the process of acquiring and effectively utilizing the knowledge, attitudes, and skills necessary to understand and regulate emotions, make responsible decisions, and maintain positive social relationships (Elias & Zins, 2006; Ross & Tolan, 2018). Emotion regulation, a critical aspect of SEL, refers to the strategies and processes that individuals use to manage their emotional experiences effectively [17]. Research has consistently shown that strong emotion regulation skills are positively associated with social competence and can serve as predictors of future social functioning [31]. In collaborative learning environments, these skills are essential for managing group dynamics, maintaining positive interactions, and fostering a sense of school belonging, particularly for students with specific learning disorders. Effective emotion regulation can lead to fewer psychosocial difficulties, contributing to a more inclusive and supportive learning environment [32], [33].

The importance of integrating SEL and emotion regulation into educational curricula cannot be overstated. Emotion regulation significantly impacts mental health by shaping emotional responses and overall psychological well-being, particularly during adolescence critical period for cognitive, social, and emotional development. Recent trends indicate that adolescents are increasingly struggling to manage their emotions effectively, which raises concerns about the potential consequences on their mental health and academic performance. Educators play a pivotal role in supporting the development of these skills through various approaches, including experiential learning, mindfulness, and simulation-based exercises [34]. Frameworks such as the RULER approach and Lobczowski's model provide structured methodologies for implementing SEL and understanding emotion formation in collaborative settings [28], [35]. By embedding emotion regulation strategies into daily educational practices, educators can enhance both their own and their students' emotional intelligence, ultimately leading to improved academic outcomes and overall well-being.

D. The Role of Virtual Reality in Enhancing Social-Emotional Learning

Cultural VR has emerged as a promising tool for enhancing SEL in children and adolescents, providing immersive environments that facilitate key skills such as perspective-taking, empathy, and responsible decision-making. VR interventions have been proven effective in enhancing SEL, especially for children with autism spectrum disorders (ASD). ASD often involve challenges with social interactions, communication, and understanding social cues. VR offers a unique, controlled, and immersive environment where children with ASD can practice and develop these essential social skills in a safe, structured setting. The studies have demonstrated the effectiveness of VR interventions in promoting SEL, particularly for children with autism spectrum disorders, where it aids in the development of social skills [36], [37]. For instance, VR can simulate real-life social situations that children with ASD might find challenging, allowing them to rehearse and refine their responses in a way that is both engaging and low-risk.

These simulations can include scenarios like having a conversation, recognizing emotions, or navigating social events. By repeatedly practicing these scenarios in a VR environment, children with ASD can build confidence and competence in their social interactions, which can translate to improved social skills in real-world settings. By offering experiential learning opportunities, VR enhances student engagement and learning outcomes across various domains, including emotional regulation, cognitive skills, and social competencies (Kamei & Harriott, 2020; Murphy et al., 2021). It shows how VR's ability to create these dynamic, engaging learning experiences leads to better retention and application of skills, making it a powerful tool in education. Furthermore, the integration of VR and augmented reality tools has shown potential in developing emotional intelligence skills, benefiting both neurotypical individuals and those with autism [40]. These findings underscore VR's role as an effective pedagogical tool for promoting SEL in diverse educational settings.

E. Comparison with Other Methods

Table I presents a comparison with various models, demonstrating that the VRSEL framework excels due to its comprehensive approach to improving emotion regulation in adolescents leveraging VR technology. Through integrating immersive experiences with tailored learning pathways and secure practice environments, it mitigates numerous constraints linked to conventional educational approaches, including restricted participation and passive learning.

TABLE I. COMPARISON WITH OTHER METHODS

Methods	Advantages	Limitations
Traditional Learning	Familiarity, structured curriculum	Limited engagement, passive learning
Online Learning Platforms	Flexibility, accessibility	Lack of immersive experience, potential isolation
Role-Playing Activities	Interactive engagement, real-world application	May not fully simulate emotional scenarios
VRSEL Framework	Immersive, personalized, risk-free practice, enhanced collaboration	Requires access to technology, potential initial learning curve

Building on this foundation, immersive technologies like VR extend their potential beyond SEL to also play a significant role in emotion regulation and the induction of positive emotions. The immersive nature of VR enhances emotional experiences by increasing the sense of presence, which can boost positive emotions and arousal [41]. Moreover, VR scenarios that offer virtual social support have been found to effectively mitigate negative emotions caused by social exclusion [42]. These applications of VR in emotion regulation span various contexts, from clinical interventions to mood induction, providing dynamic opportunities to assess and intervene in emotional processes [43], [44]. While VR holds promise for facilitating transitions from negative to positive emotional states, challenges remain in accurately representing emotions and providing design guidance [45]. Future research should focus on leveraging VR's capabilities to enhance social interactions and couple experiential training with reflective moments, further expanding its applications in emotion regulation contexts.

Despite significant advancements in VR design frameworks aimed at improving usability, user experience, and effectiveness across various domains, a critical gap exists in their application for SEL and emotion regulation. While VR has shown promise in enhancing social skills and emotional intelligence, especially for individuals with autism spectrum disorders, there is limited research on how to effectively design VR environments specifically for emotion regulation. Addressing this gap is crucial as emotion regulation is a vital component of mental health and well-being, particularly in adolescents who are increasingly struggling with managing their emotions. Understanding how VR can be optimized to support emotion regulation could lead to the development of powerful tools that not only enhance SEL but also contribute to better mental health outcomes. This study is important because it seeks to create a VR design framework that specifically targets emotion regulation, offering a new avenue for educational and therapeutic interventions in this critical area.

III. METHODOLOGY

Fig. 1 shows the methodology flowchart illustrating the study's complete process, which includes six steps in the development of the VRSEL framework.

A. Research Deign

1) Objective: The study aims to develop a specialized virtual reality design framework for SEL with the goal of enhancing emotion regulation among adolescents. As part of this study, interviews were conducted to explore and understand the virtual reality components essential for creating a design framework tailored to SEL contexts.

2) *Qualitative approach:* The study utilizes qualitative methodologies, emphasizing the in-depth data gathered through interviews. The interview questions were developed based on the initial data collected during the first phase of the research, which helped shape the conceptual framework of the study.

Fig. 1. Methodology flowchart.

B. Ethical Consideration

1) Ethical approval: This study has earned over the Research Ethics Committee's approval, The National University of Malaysia with the ethics approval number, UKM PPI/111/8/JEP-2023-111. The Malaysian Ministry of Education has also given its permission (KPM.600-3/2/3-eras (15996)).

2) Confidentiality: We maintained the participant's confidentiality by anonymizing the data. All data collected during the study was securely stored and maintained under strict confidentiality for the entire duration of the research.

C. Participant

1) Data sampling: The data sampling involves six participants from two distinct domains of expertise: VR and SEL. Participants were selected based on their specialized knowledge relevant to the study.

2) Participant selection criteria: The selection of participants was based on their specialized knowledge in two critical areas namely, virtual reality (VR) and social-Emotional learning (SEL). This dual expertise is essential for providing a comprehensive evaluation of the VRSEL framework. Participants were chosen for their ability to contribute insights that align with the study's objectives:

a) VR experts: Individuals with a strong background in VR technology, design, and implementation. Their insights are crucial for understanding the technical aspects of the VRSEL framework and its effectiveness in creating immersive learning experiences.

b) SEL experts: Professionals with expertise in social-emotional learning theories, practices, and interventions. Their perspectives are vital for evaluating the framework's alignment with established SEL principles and its impact on emotional regulation.

3) Demographics: The six participants ranged in age from 30 to 65 years. The participant group included both men and women, ensuring a balanced representation that can contribute to a more comprehensive understanding of the VRSEL framework's impact across different gender perspectives.

4) Informed Consent: Informed consent was obtained from all participants, ensuring that the study adhered to ethical standards.

D. Data Collection

1) Interview procedure: We conducted the semi-structured interview both face-to-face and online.

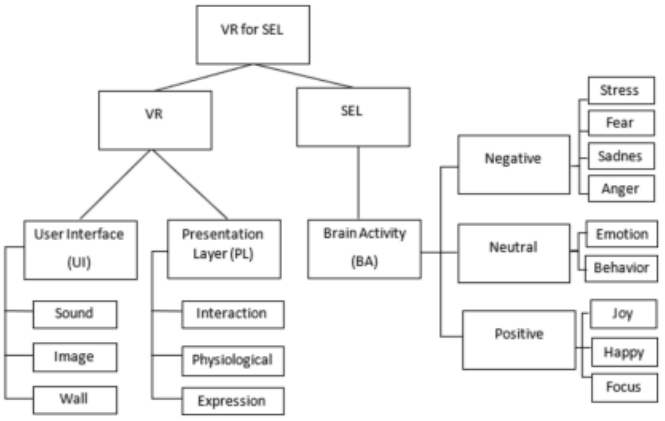


Fig. 2. Conceptual framework of VRSEL.

2) Interview guide: The preliminary data from phase one of the study, which established the study's conceptual framework [46], served as the basis for designing the interview question. Hamzah stated that it is important to have a set of requirements for making a virtual reality for social and emotional learning that could help Malaysian teens better control their feelings. Fig. 2 shows the conceptual framework of virtual reality design framework that has been designed for social-emotional knowledge developed in the earlier phase of the study by [46].

3) Recording and transcription: We recorded the interview using a voice recorder. The interview data were then transcribed using software by Google Cloud, speech-to-text, to ensure accuracy in capturing participants' responses.

E. Data Analysis

The data is analyse using thematic analysis. Thematic analysis (TA) is an adaptable method for analysing qualitative data and patterns to identify the underlying meaning. Braun and Clarke state that thematic analysis is not constrained by any one theoretical framework, rendering it a flexible instrument in qualitative research [21]. Fig. 3 displays the initial thematic map that has been produced through the process of thematic analysis. The process of TA typically involves several key steps:

1) Familiarization: The interview transcriptions were subjected to thematic analysis to ensure thorough familiarization with the data through repeated reading and examination of the transcripts.

2) Initial coding: The procedure of deriving initial codes from the data has been completed. The codes are essential since they establish the groundwork for themes.

3) Theme development: The study grouped codes into potential themes as a way of summarizing sections of the data. This involves the iterative process of reviewing themes, refining them, and ensuring they accurately represent the data. Fig. 3 represents the initial thematic map that we create at the start of data processing. The map was subsequently utilized in the development of the VRSEL framework. The theme was further examined in the discussion section of the study by providing further details on the quotation from the interview dialogue.

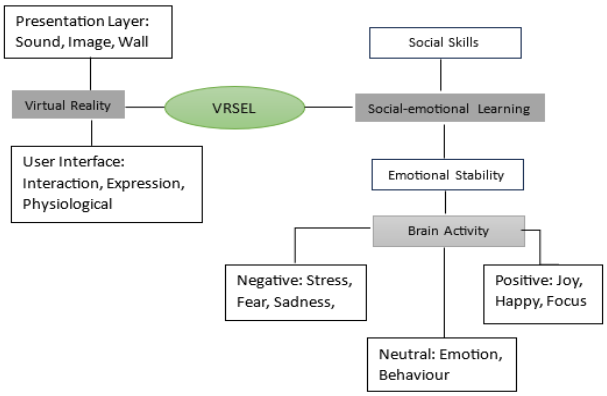


Fig. 3. Initial thematic map.

4) Reviewing themes: The themes were cross-referenced with the dataset to verify their coherence and significance.

5) Defining and naming themes: Table II indicates the definition and naming of the themes.

TABLE II. THEME DEFINITION

Themes	Sub Themes	Attributes
Virtual Reality (VR)	User Interface (UI)	Sound
		Image
		Wall
	Presentation Layer (PL)	Interaction
		Physiological
		Expression
Social-emotional Learning (SEL)	Social Skills (SS)	
	Emotional Stability (ES) Brain Activity	Stress
		Fear
		Sadness
		Anger
		Emotion
		Behavior
		Joy
		Happy
		Focus

Each theme has been delineated based on the objectives of the investigation. Subsequently, we selected the suitable titles that effectively capture the fundamental nature of the subjects. The primary focal points revolved around virtual reality and social-emotional development.

IV. FINDINGS

Fig. 4 illustrates the process of developing the VRSEL framework. The VRSEL framework is a structured approach to research that consists of four distinct stages:

Fig. 4. VRSEL framework development process.

A. Critical Literature Revie

The first stage is conducting a comprehensive examination of the current literature that is pertinent to the study area. A preliminary literature review was conducted in the initial phase of the project [46]. The study highlighted any deficiencies in knowledge, comprehending the present state of VR framework for SEL, and providing a rationale for the necessity of the study. It provides information on the research questions and objectives, which helps to drive the overall progress of the project.

B. User Requirement Data

During the second phase, data has been collected pertaining to user requirements. This entails identifying the requirements, interests, and expectations pertaining to the extent of the investigation. Method such as interviews have been utilized to gather qualitative data. This phase is crucial for guaranteeing that the framework being constructed is focused on the user and effectively tackles current issues. The utilization of theme analysis in this data analysis yields valuable insights that may be employed to tailor the research to meet the specific needs of users, hence enhancing its relevance and efficacy.

C. Framework Design

The third phase is dedicated to the development of the framework's design. This entails integrating the knowledge acquired from the literature review and user requirement data to develop a logical framework that directs the study. The framework has explicitly outlined the interconnections among different components of the study, including variables, themes, and attributes as illustrated in the Fig. 5.

1) Key themes identified: Several significant topics pertinent to the VRSEL paradigm emerged through thematic analysis:

a) Theme 1: Virtual Reality (VR): It includes subtheme of user interface (UI) and presentation layer (PL).

b) Theme 2: Social-Emotional Learning (SEL): It includes subtheme of social skills (SS) and emotional stability (ES).

D. Testing and Validation

The final phase involves the validation of the VRSEL framework and assessing the usability of the VRSEL framework in the next study. This is accomplished by utilizing experimental approaches to evaluate the efficacy and dependability of the framework for future investigations. Validation and usability testing may include pilot studies, expert review, user feedback, and VR app development based on the results obtained. Continuous testing and validation help in refining the framework, making it more robust and applicable to real-world scenarios.

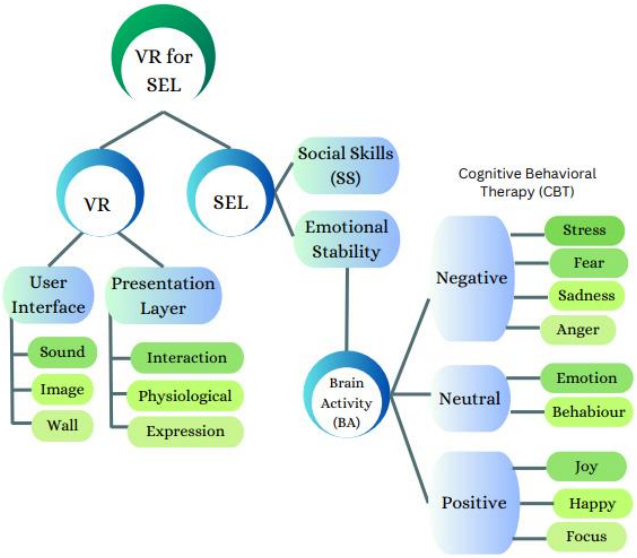


Fig. 5. VRSEL framework.

Fig. 5 highlights the primary findings of the study, which revolve around the VRSEL framework. The VRSEL framework encompasses two primary themes, specifically virtual reality (VR) and social-emotional learning (SEL). The VR theme comprises two subthemes such as user interface (UI) and presentation layer (PL). The UI features include sound, image, and wall. The characteristics of the PL encompass interaction, physiological responses, and expression. Meanwhile, the SEL theme comprises two subthemes namely, social skills (SS) and emotional stability (ES). The ES has a sub-subtheme called brain activity (BA). The BA possesses traits that can be categorized as negative, neutral, or good. The subtheme of ES is integrated with cognitive behavioural therapy (CBT).

V. DISCUSSION

A. Design Principles of the VRSEL Framework

The findings indicate that the VRSEL framework offers the necessary components for creating a targeted VR design within the context of SEL, as compared to the prior study on design framework. Table III displays the code that was produced during

the process of thematic analysis, which included interviews with all participants. Participant 3 highlight that "Social emotional development is divided into two, namely, social skills and emotional stability."

TABLE III. VRSEL INTERVIEW CODE

Codes	Interview Clip
Sound	"We need to place the sound according to the suitability of our VR application"
Image	"The objective of these computerized graphics is to get a greater impact through visuals"
Wall	"So apart from the wall, why don't we use virtual reality to visually represent avatars and other individuals?"
Interaction	"Interactive VR is one of the most important elements to get input from the user"
Physiological	"VR requires physiological movement"
Expression	"You are the virtual storyteller. You represent your own world."
Social Skill	"Social skill is an ecologically valid social skill step."
Emotional Stability	"Emotional stability is needed for someone to learn about emotions"

The VRSEL interview codes play a role in advancing the VRSEL framework. The finding highlights four fundamental design principles for its development:

1) User Interface (UI): Interaction design principles are a set of standards that assist in the creation of user interfaces that are both intuitive and engaging. These principles aim to replicate real-world social circumstances, resulting in lifelike scenarios. Sound is essential in constructing engaging and credible virtual reality settings. Sound contributes to the establishment of a feeling of being present and fully engaged in the virtual environment. The use of realistic and spatially precise audio enhances the authenticity and persuasiveness of the VR experience, immersing the user more profoundly into the scenario [47]. Creating a VR scenario needs to incorporate a diverse range of visual components, such as 360-degree image, 3D models, textured backgrounds, lighting, and interactive features [47]. Moreover, an image illustrating the concept of virtual reality, where a virtual environment may be explored using an avatar. In addition to using a virtual wall, we can enhance the immersion of virtual reality by incorporating an avatar. An avatar serves as the user's representative in the virtual world, allowing them to interact with the environment [2]. Avatars serve as digital embodiments of users, playing a vital role in creating a feeling of being present in the virtual world. Through interacting with their avatars, users could manifest their unique characteristics and personal sense of self, thereby intensifying their involvement and emotional connection to the overall experience [48].

2) Presentation Layer (PL): The presentation layer (PL) is an essential element in the design of VR, specifically in relation to interaction, physiological aspects, and user expression. It enables users to engage with virtual characters and situations to develop SEL capacities. The PL is accountable for overseeing user interactions within the VR environment. VR experiences can induce notable physiological effects on users, including symptoms such as motion sickness, eye strain, and weariness. The PL must be carefully crafted, taking into consideration these considerations, to ensure that visual elements are optimized to reduce any discomfort. The PL in social VR experiences must facilitate user expression through avatars and other virtual representations. These aspects must precisely communicate the user's movements, gestures, and facial expressions, facilitating authentic and captivating interactions with other users.

3) Social Skills (SS): Social skills are a crucial component of SEL, encompassing the abilities necessary for effective communication and building connections with others. These skills include understanding and empathizing with others, resolving conflicts in a positive manner, and demonstrating cultural humility and respect for diversity.

4) Emotional Stability (ES): Emotional stability (ES), as defined by SEL, is the capacity to skilfully control and regulate one's emotions, ensuring a harmonious emotional state in the face of diverse circumstances. This concept is closely linked to brain activity (BA) and how individuals respond to negative, neutral, and positive stimuli. BA that encompasses responses such as negative, neutral, and positive provides prompt feedback and a chance for introspection to enhance learning and incorporate cognitive brain therapy (CBT).

In Table II, SEL theme demonstrates the display of negative emotions such as stress fear, sadness, or anger. When exposed to negative stimuli, the amygdala in the brain becomes activated, resulting in intensified emotional reactions such as anxiety or wrath. Individuals who possess ES could effectively handle these responses, using techniques acquired via SEL to reduce the influence of negative feelings. Neutral stimuli, such as emotions and behaviours, typically do not elicit intense emotional reactions, enabling humans to analyse information without any prejudice. Emotional stability allows an individual to sustain concentration and lucidity in neutral circumstances, which is crucial for making decisions and resolving problems. Positive stimuli elicit the release of neurotransmitters, such as dopamine, which stimulate feelings of enjoyment and motivation such as joy happy and focus that is illustrated in Fig. 5. Individuals who possess ES can completely participate in pleasant events, which in turn improves their emotional well-being and strengthens their ability to adjust to different situations.

VI. FUTURE WORK

The results suggest that the VRSEL framework has significant potential to facilitate SEL by providing immersive and interactive experiences that foster emotional and social development. However, it is crucial to conduct thorough testing and validation of the VRSEL framework. A forthcoming phase will involve integrating the VRSEL framework into the development of a VR application for rigorous testing and validation. Implementing VR for SEL presents several challenges, including technical limitations, content development complexities, and the necessity for specialized training. From a technical perspective, VR systems demand high-performance hardware and software to deliver immersive experiences, which

can be costly and challenging to maintain, particularly in educational settings with limited resources. Additionally, creating content that is both engaging and pedagogically sound requires substantial expertise in VR technology and SEL principles. Developing realistic, interactive scenarios that effectively teach social-emotional skills necessitates close collaboration among educators, VR developers, and psychologists. Furthermore, educators and facilitators need comprehensive training not only to use VR tools but also to integrate them meaningfully into SEL curricula. This often involves a steep learning curve and ongoing support, which can hinder widespread adoption. To address these challenges, schools and institutions might consider forming partnerships with VR developers, investing in scalable training programs, and exploring cost-effective VR solutions that still offer immersive learning experiences. Additionally, ongoing research and feedback loops are essential to refining VR content and enhancing its accessibility and effectiveness across diverse educational environments.

VII. Conclusion

The VRSEL framework marks a significant advancement in integrating technology with educational practices, particularly in fostering SEL. By harnessing the immersive nature of VR, this framework creates interactive, lifelike scenarios where learners can actively engage in experiences that cultivate emotional intelligence, empathy, and other essential SEL skills among adolescents especially in Malaysia. Through virtual environments that replicate real-world social situations, VRSEL allows students to practice and enhance their emotional responses, decision-making, and interpersonal interactions in a safe, controlled setting. This experiential approach offers a deeper understanding and retention of SEL concepts compared to traditional methods.

However, the successful implementation of VRSEL in educational settings necessitates careful design and consideration of various factors. The framework's effectiveness hinges on the alignment of virtual experiences with learning objectives and the seamless integration of the technology into existing curricula. Designing these virtual environments requires a solid foundation in SEL principles and must be tailored to the diverse needs of learners, including varying learning styles and abilities. Additionally, ensuring that the content is culturally relevant and inclusive is vital for creating a positive learning experience for all students. Continued research and development are essential to fully realize the VRSEL framework's potential. While the benefits are evident, challenges such as technical limitations, content development, and the need for comprehensive educator training must be addressed. Ongoing studies are necessary to refine the framework, optimize the user experience, and assess the long-term impact of VRSEL on student outcomes. By overcoming these challenges and fostering innovation, VRSEL has the potential to transform SEL education, making it more engaging, effective, and accessible to learners worldwide.

Acknowledgment

I. Hamzah and E. Salwana expresses gratitude to the Ministry of Education, Malaysia for their crucial assistance in funding this work, provided under the Trans-disciplinary Research Grant Scheme under the code of TRGS/1/2020/UKM/01/4/3.

References

[1] S. Mandal, "Brief Introduction of Virtual Reality & its Challenges," vol. 4, no. 4, pp. 304–309, 2013.

[2] C. Girvan, "What is a virtual world? Definition and classification," Educational Technology Research and Development, vol. 66, no. 5, pp. 1087–1100, 2018, doi: 10.1007/s11423-018-9577-y.

[3] I. Wohlgenannt, A. Simons, and S. Stieglitz, "Virtual Reality," Business & Information Systems Engineering, 2020, doi: 10.1007/s12599-020-00658-9.

[4] J. Radianti, T. A. Majchrzak, J. Fromm, and I. Wohlgenannt, "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda," Comput Educ, vol. 147, no. November 2019, p. 103778, 2020, doi: 10.1016/j.compedu.2019.103778.

[5] S. Partelow, "What is a framework? Understanding their purpose, value, development and use," J Environ Stud Sci, vol. 13, no. 3, pp. 510–519, Sep. 2023, doi: 10.1007/s13412-023-00833-w.

[6] J. C. Moullin et al., "Ten recommendations for using implementation frameworks in research and practice," Implement Sci Commun, vol. 1, no. 1, Dec. 2020, doi: 10.1186/s43058-020-00023-7.

[7] M. H. Hatta et al., "Virtual Reality (VR) Technology for Treatment of Mental Health Problems during COVID-19 : A Systematic Review," 2022.

[8] A. Christopoulos, N. Pellas, and M. J. Laakso, "A learning analytics theoretical framework for stem education virtual reality applications," Educ Sci (Basel), vol. 10, no. 11, pp. 1–15, Nov. 2020, doi: 10.3390/educsci10110317.

[9] A. L. Green, S. Ferrante, T. L. Boaz, K. Kutash, and B. Wheeldon-Reece, "Social and emotional learning during early adolescence: Effectiveness of a classroom-based SEL program for middle school students," Psychol Sch, vol. 58, no. 6, pp. 1056–1069, Jun. 2021, doi: 10.1002/pits.22487.

[10] D. Yaeger, "Social and Emotional Learning Programs for Adolescents," Future of Children, vol. 27, no. 1, pp. 73–94, 2017, [Online]. Available: http://web.b.ebscohost.com.proxygw.wrlc.org/ehost/detail/detail?vid=0&sid=2b78f5d2-09a2-45cc-a273-ae3d107fc5cb%40sessionmgr103&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3D%3D#AN=123568102&db=sih

[11] J. E. Zins and M. J. Elias, "Social and emotional learning: Promoting the development of all students," Journal of Educational and Psychological Consultation, vol. 17, no. 2–3, pp. 233–255, 2007, doi: 10.1080/10474410701413152.

[12] K. M. Ross and P. Tolan, "Social and Emotional Learning in Adolescence: Testing the CASEL Model in a Normative Sample," Journal of Early Adolescence, vol. 38, no. 8, pp. 1170–1199, Oct. 2018, doi: 10.1177/0272431617725198.

[13] G. M. Lawson, M. E. McKenzie, K. D. Becker, L. Selby, and S. A. Hoover, "The Core Components of Evidence-Based Social Emotional Learning Programs," Prevention Science, vol. 20, no. 4, pp. 457–467, May 2019, doi: 10.1007/s11121-018-0953-y.

[14] T. English and L. Eldesouky, "We're not alone: Understanding the social consequences of intrinsic emotion regulation," American Psychological Association, 2020.

[15] M. W. Southward, E. M. Altenburger, S. A. Moss, D. R. Cregg, and J. S. Cheavens, "Flexible, yet firm: A model of healthy emotion regulation," J Soc Clin Psychol, vol. 37, no. 4, pp. 231–251, 2018, doi: 10.1521/jscp.2018.37.4.231.

[16] E. S. Blanke, J. A. Bellingtier, M. Riediger, and A. Brose, "When and How to Regulate: Everyday Emotion-Regulation Strategy Use and Stressor Intensity," Affect Sci, vol. 3, no. 1, pp. 81–92, Mar. 2022, doi: 10.1007/s42761-021-00087-1.

[17] J. J. Gross, "Emotion Regulation: Current Status and Future Prospects," Psychol Inq, vol. 26, no. 1, pp. 1–26, Jan. 2015, doi: 10.1080/1047840X.2014.940781.

[18] Y. Shi, P. Koval, V. Kostakos, J. Goncalves, and G. Wadley, "'Instant Happiness': Smartphones as tools for everyday emotion regulation,"

International Journal of Human Computer Studies, vol. 170, Feb. 2023, doi: 10.1016/j.ijhcs.2022.102958.

[19] N. Baghaei, V. Chitale, A. Hlasnik, L. Stemmet, H.-N. Liang, and R. Porter, "Virtual Reality for Supporting the Treatment of Depression and Anxiety: Scoping Review," JMIR Ment Health, vol. 8, no. 9, p. e29681, Sep. 2021, doi: 10.2196/29681.

[20] M. Mulders, J. Buchner, and M. Kerres, "A Framework for the Use of Immersive Virtual Reality in Learning Environments," International Journal of Emerging Technologies in Learning, vol. 15, no. 24, pp. 208–224, 2020, doi: 10.3991/ijet.v15i24.16615.

[21] V. Braun and V. Clarke, "Using thematic analysis in psychology," Qual Res Psychol, vol. 3, no. 2, pp. 77–101, 2006, doi: 10.1191/1478088706qp063oa.

[22] C. Tuena et al., "Usability issues of clinical and research applications of virtual reality in older people: A systematic review," Front Hum Neurosci, vol. 14, 2020, doi: 10.3389/fnhum.2020.00093.

[23] H. T. Chong, C. K. Lim, and K. L. Tan, "Users Experience with VR System: Current State and Development Directions," TEST Engineering & Management, vol. 82, no. January-February 2020, pp. 6429–6436, 2020, [Online]. Available: http://scholar.google.com/scholar_url?url=https://iceiee.org/index.php/testmagzine/article/download/1860/1670&hl=en&sa=X&d=13959535640894808774&scisig=AAGBfm1KVAzaRFlNHMhtlJyBscY45wPNhQ&nossl=1&oi=scholaralrt&hist=IaPXgqsAAAAJ:11798333417038799993:AAGBfm0

[24] D. Rodríguez-Almagro, A. Achalandabaso-Ochoa, A. J. Ibáñez-Vera, J. Góngora-Rodríguez, and M. Rodríguez-Huguet, "Effectiveness of Virtual Reality Therapy on Balance and Gait in the Elderly: A Systematic Review," Jan. 01, 2024, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/healthcare12020158.

[25] Q. Lyu, K. Watanabe, H. Umemura, and A. Murai, "Design-thinking skill enhancement in virtual reality: A literature study," 2023, Frontiers Media S.A. doi: 10.3389/frvir.2023.1137293.

[26] G. Wadley et al., "The Future of Emotion in Human-Computer Interaction," in Conference on Human Factors in Computing Systems - Proceedings, Association for Computing Machinery, Apr. 2022. doi: 10.1145/3491101.3503729.

[27] G. Kaur, "Research Paper on Human Computer Interaction (HCI)," 2023. [Online]. Available: www.ijfmr.com

[28] J. D. , B. M. A. , B. C. S. , & W. C. J. Hoffmann, "Teaching emotion regulation in schools: Translating research into practice with the RULER approach to social and emotional learning. ," Emotion, vol. 20, no. 1, p. 105, 2020.

[29] E. M. Harrington, S. D. Trevino, S. Lopez, and N. R. Giuliani, "Supplemental Material for Emotion Regulation in Early Childhood: Implications for Socioemotional and Academic Components of School Readiness," Emotion, vol. 20, no. 1, p. 48, 2020, doi: 10.1037/emo0000667.supp.

[30] M. J. Elias and J. E. Zins, "Social and emotional learning 1 Social and Emotional Learning," 2006. [Online]. Available: https://www.researchgate.net/publication/284593261

[31] E. A. Matchett et al., "International Journal of Autism & Related Disabilities Emotion Regulation and Social Participation in Childhood and Adolescence: Systematic Review," 2020, doi: 10.29011/2642-3227.000038.

[32] D. Kopelman-Rubin, A. Siegel, N. Weiss, and I. Kats-Gold, "The Relationship between Emotion Regulation, School Belonging, and Psychosocial Difficulties among Adolescents with Specific Learning Disorder," Child Sch, vol. 42, no. 4, pp. 216–224, Oct. 2020, doi: 10.1093/cs/cdaa022.

[33] K. Mänty, H. Järvenoja, and T. Törmänen, "Socio-emotional interaction in collaborative learning: Combining individual emotional experiences and group-level emotion regulation," Int J Educ Res, vol. 102, Jan. 2020, doi: 10.1016/j.ijer.2020.101589.

[34] K. M. Sewell, "Examining the Place of Emotions, Affect, and Regulation in Social Work Education," J Soc Work Educ, vol. 56, no. 1, pp. 5–16, Jan. 2020, doi: 10.1080/10437797.2019.1627262.

[35] N. G. Lobczowski, "Bridging gaps and moving forward: Building a new model for socioemotional formation and regulation," Educ Psychol, vol. 55, no. 2, pp. 53–68, Apr. 2020, doi: 10.1080/00461520.2019.1670064.

[36] A. Frolli et al., "Children on the Autism Spectrum and the Use of Virtual Reality for Supporting Social Skills," Children, vol. 9, no. 2, Feb. 2022, doi: 10.3390/children9020181.

[37] F. Ke, J. Moon, and Z. Sokolikj, "Virtual Reality–Based Social Skills Training for Children With Autism Spectrum Disorder," Journal of Special Education Technology, vol. 37, no. 1, pp. 49–62, Mar. 2022, doi: 10.1177/0162643420945603.

[38] A. Kamei and W. Harriott, "Social emotional learning in virtual settings: Intervention strategies," International Electronic Journal of Elementary Education, vol. 13, no. 3, pp. 365–371, 2020, doi: 10.26822/IEJEE.2021.196.

[39] K. M. Murphy, A. L. Cook, and L. M. Fallon, "Mixed reality simulations for social-emotional learning," vol. 102, no. 6, pp. 30–37, 2021.

[40] C. Papoutsi, A. Drigas, and C. Skianis, "Virtual and Augmented Reality for Developing Emotional Intelligence Skills," International Journal of Recent Contributions from Engineering, Science & IT (iJES), vol. 9, no. 3, p. 35, Sep. 2021, doi: 10.3991/ijes.v9i3.23939.

[41] K. Pavic, L. Chaby, T. Gricourt, and D. Vergilino-Perez, "Feeling Virtually Present Makes Me Happier: The Influence of Immersion, Sense of Presence, and Video Contents on Positive Emotion Induction," Cyberpsychol Behav Soc Netw, vol. 26, no. 4, pp. 238–245, Apr. 2023, doi: 10.1089/cyber.2022.0245.

[42] L. Stallmann, M. Tran, D. Rudrauf, D. Dukes, and A. C. Samson, "Simulating Social Emotion Regulation in Virtual Reality," International Journal of Emerging Technologies in Learning, vol. 18, no. 7, pp. 4–27, 2023, doi: 10.3991/ijet.v18i07.29419.

[43] A. H. Bettis, T. A. Burke, J. Nesi, and R. T. Liu, "Digital Technologies for Emotion-Regulation Assessment and Intervention: A Conceptual Review," Jan. 01, 2022, SAGE Publications Inc. doi: 10.1177/21677026211011982.

[44] S. Vita, C. Morra, and A. Rega, "Virtual reality and emotion regulation: a systematic review," PSYCHOBIT, 2021.

[45] O. Sadka and A. Antle, "Interactive technologies for emotion-regulation training: Opportunities and challenges," in Conference on Human Factors in Computing Systems - Proceedings, Association for Computing Machinery, Apr. 2020. doi: 10.1145/3334480.3382894.

[46] I. Hamzah et al., "Virtual Reality for Social-Emotional Learning: A Review," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Science and Business Media Deutschland GmbH, 2024, pp. 119–130. doi: 10.1007/978-981-99-7339-2_11.

[47] I. de V. Bosman, O. 'Oz' Buruk, K. Jørgensen, and J. Hamari, "The effect of audio on the experience in virtual reality: a scoping review," Behaviour and Information Technology, vol. 43, no. 1, pp. 165–199, 2024, doi: 10.1080/0144929X.2022.2158371.

[48] R. Radiah, D. Roth, F. Alt, and Y. Abdelrahman, "The Influence of Avatar Personalization on Emotions in VR," Multimodal Technologies and Interaction, vol. 7, no. 4, Apr. 2023, doi: 10.3390/mti7040038.

Bridging the Gap: Machine Learning and Vision Neural Networks in Autonomous Vehicles for the Aging Population

Shengsheng Tan
School of Science, Edith Cowan University, Joondalup, Australia

***Abstract*—As autonomous vehicles (AVs) evolve recently, it is necessary to address the unique needs of the aging population group. They can get a significant benefit from this technology. This scoping review focus the role of machine learning and vision neural networks in autonomous vehicles. A focus on enhancing safety, usability, and trust for elderly users will be mentioned as well. It systematically reviews existing literature to identify how these technologies address the cognitive and physical challenges faced by older adults. The review highlights key advancements in AV technology, such as adaptive interfaces and assistive features. That can enhance the driving experience for the elderly. Additionally, it investigates factors influencing trust and acceptance of AVs among older adults, emphasizing the importance of transparent and user-friendly design. Although, the despite notable progress has been made, the significant gaps remain in understanding how to optimize these technologies to meet the diverse needs of elderly passengers. The review identifies areas for future research, including personalized AV systems and regulatory frameworks that support designs friendly to the elderly. By addressing these gaps, the study aims to contribute to developing autonomous vehicles that are inclusive and accessible. It will make the mobility and quality of life for the aging population increased. This review underscores the importance of integrating machine learning and vision neural networks in designing AVs that cater to the unique needs of older adults. It was also offering valuable insights for researchers, policymakers, and industry stakeholders advancing autonomous vehicle technology.**

Keywords*—Autonomous vehicles; machine learning; vision neural network; human-computer interaction; aging population; artificial intelligence

I. Introduction

A. Overview of Autonomous Vehicles (AVs) and their Increasing Role in Transportation

Autonomous vehicles (AVs), also known as self-driving cars. It's capable of operating with reduced or no human input. AVs are making a revolution of the transportation landscape by offering new possibilities for mobility and accessibility [1]. These vehicles utilize a combination of sensors, cameras, and algorithms to navigate roads without human intervention [2]. The progress in vehicles' technology have accelerated the development and deployment of AVs over the past decade. The industry has positioned them as a transformative force in the automotive industry [3]. The promise of enhanced safety, efficiency, and convenience has fueled interest in AVs from governments, industry stakeholders, and consumers alike [4]. AVs offer a potential solution to reduce traffic problems as urbanization and population growth place increasing demands on transportation infrastructure. For instance, improve road safety and provide accessible transportation options for diverse populations [3].

B. Introduction to Machine Learning and Vision Neural Networks in AVs

Machine learning and vision neural networks are key technologies driving the evolution of autonomous vehicles [5]. Machine learning algorithms make AVs to learn from large datasets. This allowing them to recognize patterns, make decisions, and adapt to changing environments in different driving scenarios [6]. Vision neural networks is a subset of machine learning. They are particularly crucial for AVs as they process and interpret visual information from cameras and sensors. These networks enable AVs to detect and identify objects, interpret traffic signals, and navigate complex environments [7]. Technologies like Tesla's Full Self-Driving (FSD) system exemplify the application of vision neural networks in AVs, allowing for advanced features such as automated lane changes and obstacle avoidance [8]. The integration of these technologies enhances the capability of AVs in real-world conditions, so AVs can operate safely and efficiently on the road.

C. Importance of Addressing the Needs of the Aging Population in the Context of AV Technology

The aging population represents a significant and growing demographic, especially in Australia. Approximately 4.2 million people were aged 65 and over on 30 June 2020, this occupied 16% of Australia’s total population [9]. This group will stand to benefit greatly from autonomous vehicle technology. As the age growing, elderly may face cognitive and physical challenges that impact their ability to drive safely. Autonomous vehicles can provide a more mobility and independence ability to elderly individuals. It will reduce elder people's reliance on public transportation or family assistance [10]. However, it's essential to address the specific needs and preferences of older adults in the design and implementation of AV technology to fully meet these benefits. This includes considerations for cognitive load, ease of use, and safety features that cater to the physical limitations and cognitive changes associated with aging [11]. For example, AV technology can be developed to enhance their quality of life and promote social inclusion to focus on the unique requirements which elderly need. Additionally, understanding

and addressing the trust and acceptance factors among elderly users is critical for successful AV adoption. As this group of people continues to grow, incorporating their needs into AV design will improve accessibility and AVs can contribute positively to societal well-being at the same time [12].

II. METHODOLOGY

The methodology for this scoping review follows a structured approach to ensure comprehensive coverage of the literature related to machine learning and vision neural networks in autonomous vehicles (AVs) for the aging population. This section outlines the framework used for conducting the review, the search strategy employed to identify relevant studies, the criteria for inclusion and exclusion, and the process of data extraction and analysis.

A. Scoping Review Framework

This scoping review is conducted following the framework proposed by Arksey and O'Malley (2005), which is widely used in scoping reviews to map the existing literature and identify gaps in research. The framework of the scoping review includes of five key stages:

1) Identifying the research question: The primary research question guiding this review is: "What is the current stage of research on machine learning and vision neural networks in autonomous vehicles? How are they can help the aging population?"

2) Identifying relevant studies: A comprehensive search of academic databases was conducted to identify relevant literature. The studies published in the last four years (2020-2024) will be focused, aims to capture the most recent advancements in this field.

3) Selecting studies: The study selection process should followed by predefined inclusion and exclusion criteria, to ensure only the pertinent studies in the review inclusion.

4) Charting the data: Data from selected studies were extracted and charted to facilitate analysis and synthesis.

5) Collating, summarizing, and reporting the results: The results were collated, summarized, and reported to provide an overview of the current state of research and identify gaps and future research directions.

B. Search Strategy

The search strategy was designed to capture a comprehensive set of studies related to the use of machine learning and vision neural networks in AVs for elderly users. The following steps were taken to implement the search strategy:

1) Databases and sources: The literature search was conducted using the following academic databases, which are known for their extensive coverage of scientific and technical publications:

a) PubMed: For biomedical and life sciences literature.

b) IEEE Xplore: For engineering and technology-focused studies.

c) Google Scholar: For a broad search across various disciplines.

d) Scopus: For a wide range of reviewed literature in the fields of science, technology, medical, and social sciences.

2) Keywords and search terms: The search terms were selected by the keywords of this review topic to make sure it will catch the relevant literature on the topic. Keywords used in the search included: "autonomous vehicles", "self-driving cars", "machine learning", "deep learning", "vision neural networks", "elderly users", "aging population", "senior citizens", "accessibility", "trust and acceptance", "cognitive challenges", “human-computer interaction”.

3) Search syntax: Boolean operators (AND, OR) were used to combine search terms and refine the search results. For example, the search string "autonomous vehicles AND machine learning AND elderly users" was used to identify studies that specifically addressed the intersection of these topics.

4) Timeframe: The search had a time limited for each article. They must have been published between January 2020 and August 2024 to ensure the inclusion of the most recent research.

C. Inclusion and Exclusion Criteria

The inclusion and exclusion criteria were established to ensure that the studies selected for the review were relevant and of high quality. The criteria are as follows:

1) Inclusion Criteria

a) Studies that focus on autonomous vehicles and their application to the aging population.

b) Research involving machine learning and vision neural networks in the context of AVs.

c) Empirical studies, including quantitative, qualitative, and mixed-methods research.

d) Review journal articles, conference papers, and academic journals.

e) Publications in English to ensure accessibility and comprehension.

2) Exclusion Criteria

a) Studies not related to the aging population or elderly users in the context of AVs.

b) Articles that do not involve machine learning or vision neural networks.

c) Non-empirical studies, such as opinion pieces, editorials, and commentaries.

d) Publications not available in English or some lacking full-text access articles.

D. Data Extraction and Analysis

Data extraction and analysis were conducted systematically to ensure accurate and comprehensive synthesis of the literature. The following steps were undertaken:

1) Data Extraction

a) A standardized data extraction form was developed to capture key information from each selected study. The form included fields for:

i) Study title, authors, and publication year.
ii) Research objectives and questions.
iii) Methodology and study design.
iv) Sample characteristics and size.
v) Key findings and conclusions.
vi) Implications for AV design and implementation for the aging population.

2) Data synthesis

a) The extracted data were synthesized because the goal of identifying commonalities is to among themes, trends, and gaps in the literature. Thematic analysis has been done on the categorized findings around certain topics like accessibility, trust, cognitive challenges, and technology acceptance.

3) Identifying key themes and trends

a) Based on the study, the critical themes and trends in the literature are the role of machine learning in enhancing AV accessibility for the elderly, the impact that visual neural networks have on user experience, and factors influencing trust and acceptance among older adults.

4) Reporting the results

a) It shows the current research stage, research gaps, and possible directions for future studies in this area. The results will be useful to inform the design and development of AVs that are accessible and user-friendly for the aging population.

5) Quality assessment

a) Although scoping reviews generally do not involve a formal quality assessment, the studies identified were assessed with respect to methodological rigor and relevance to ensure the strength of the findings.

III. FINDINGS

A. Overview

A comprehensive search obtained by computer was conducted across in four major databases which are PubMed, IEEE Xplore, Scopus, and Google Scholar. As a result, there was a total of 5,702 articles (with 19,300 results from Google Scholar) has been yielded. Also, there were 158 articles were obtained by manual retrieval of references of included articles. A total of 732 articles were eliminated by EndNote20 software and manual deduplication. After reading the title, 4,969 articles that obviously did not meet the research theme were eliminated. After reading the abstract and combining the inclusion and exclusion criteria, 701 articles were excluded. Finally, 31 articles were included [13-43]. The literature screening process of this study is shown in Fig. 1. All selected journal articles explored the integration of machine learning and vision neural networks in autonomous vehicles (AVs) with a specific focus on the aging population. The geographical distribution of the studies included North America (8 studies) [14] [20] [23] [24] [25] [26] [30] [38], Europe (8 studies) [16] [17] [22] [29] [31] [32] [37] [41], Asia (13 studies) [15] [18] [19] [21] [28] [33] [34] [35] [36] [39] [40] [42] [43], and other regions (2 studies) [13] [27]. Please refer to Fig. 2 for details.

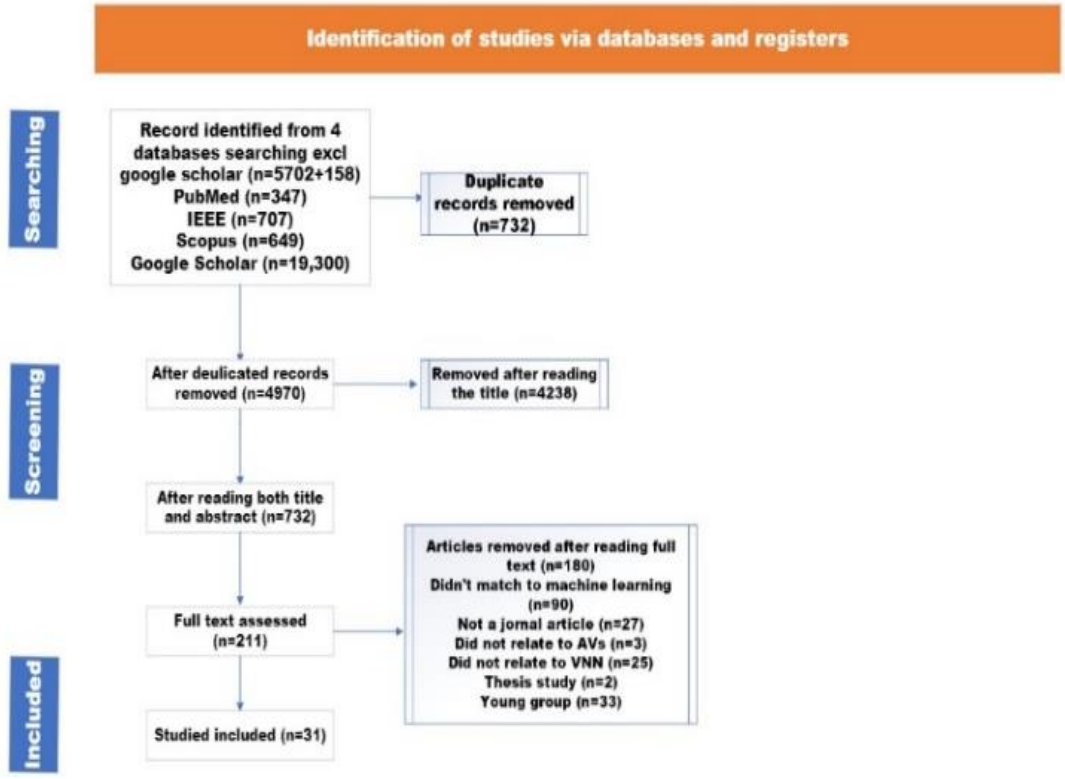


Fig. 1. Flow chart of literature screening.

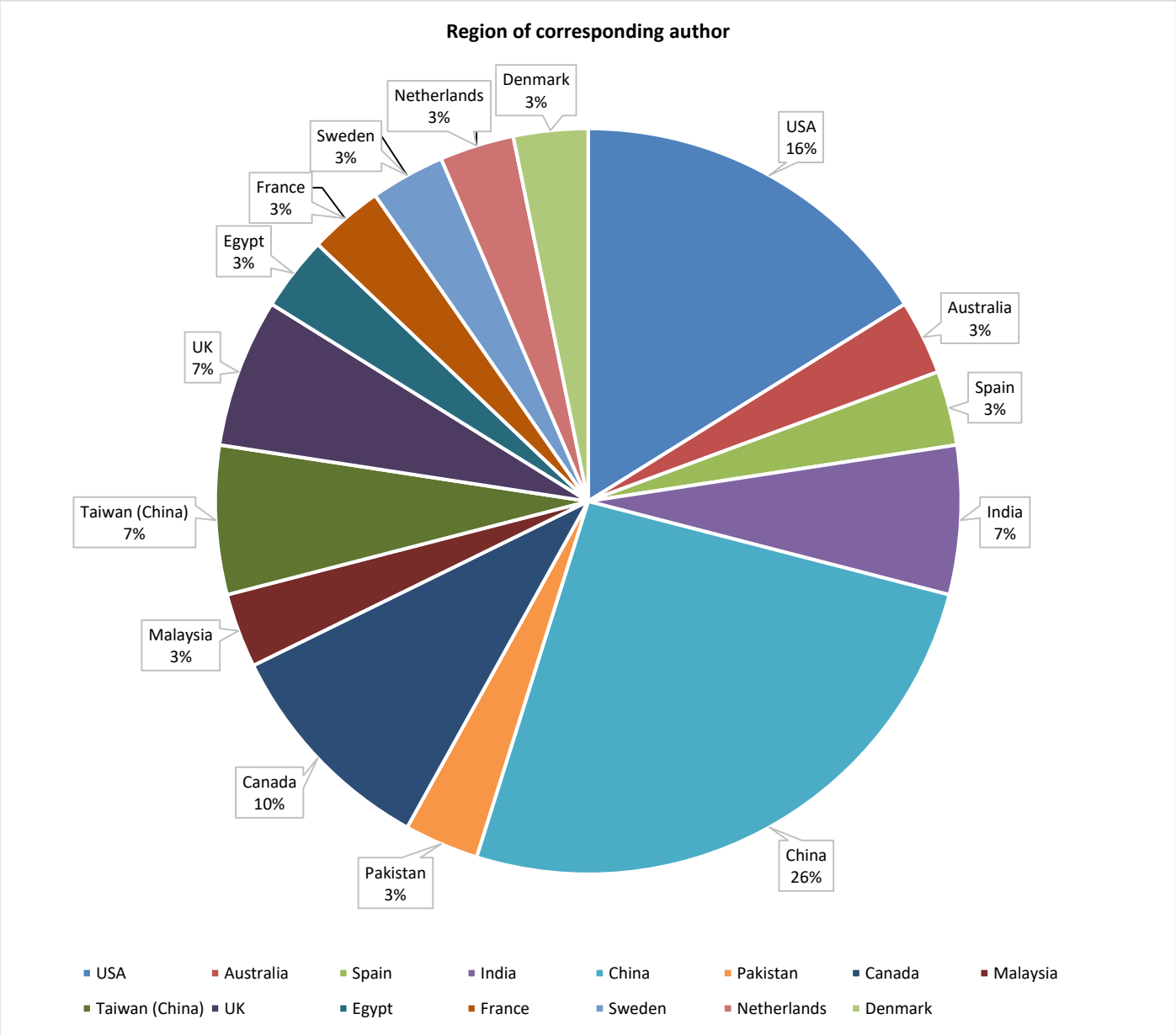


Fig. 2. Chart of articles' geographical distribution.

B. Sample Characteristics

The reviewed studies included diverse sample sizes and participant demographics. Of the 31 studies, 15 involved human participants, with a particular focus on elderly and disabled individuals [16] [19] [20] [22] [24] [25] [26] [27] [29] [31] [32] [34] [38] [42] [43]. The sample sizes in these studies in a range from 50 participants to 3,500 participants. Those with larger sample sizes were typically sourced from national health surveys and electronic health records (EHRs) which are over 1000 participants. The remaining 16 studies utilized simulation environments and predictive models to examine interactions between AV systems and various user groups. Many studies highlighted the importance of large, diverse datasets to properly train machine learning models, especially to cater to the needs of underrepresented groups.

C. Types of Machine Learning Algorithms

The studies employed various machine learning algorithms to enhance AV functionality and safety for elderly and disabled users. The most commonly used method was Tree-based methods. It's including Random Forest and Gradient Boosting Machines (GBM). There were about in 12 studies appeared [14] [29] [32] [40]. These algorithms were preferred for their robustness in managing complex datasets. There were seven studies mentioned Neural networks, this included Convolutional Neural Networks and Recurrent Neural Networks [14] [15] [32] [39] [40]. The primary use for vision-based tasks like object detection and predicting user behavior. Support Vector Machines (SVMs) and Bayesian Networks were featured in five studies [14] [19] [29] [32] [42], particularly for predicting user trust and acceptance of AV technology among elderly users (Fig. 3).

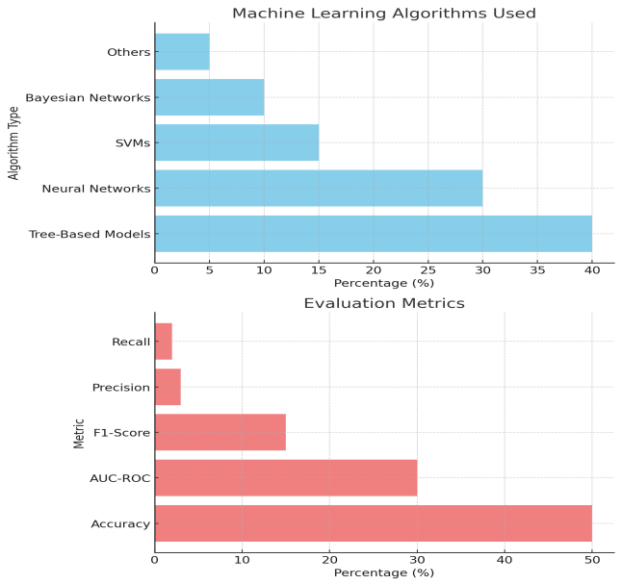


Fig. 3. ML algorithms used & evaluation metrics.

D. Handling of Missing Data

Handling missing data was a key focus in several studies due to the complexity and diversity of the datasets involved. Of the 31 studies, nine specifically discussed their strategies for dealing with missing data. Multiple imputation was the most commonly used technique, which reported in five studies. On the other hand, there were four studies used k-Nearest Neighbors. These methods were critical for maintaining dataset integrity, especially in studies involving health data or surveys where missing values were prevalent.

E. Evaluation Metrics and Model Performance

The studies used a variety of metrics to evaluate the performance of machine learning models in AV systems. The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) was the most frequently reported metric, used in 14 studies, to assess the models' ability to distinguish between successful and unsuccessful AV operations involving elderly and disabled passengers. Calibration plots and Brier scores were also used in 6 studies to evaluate the accuracy of predictive models [25]. The average AUC-ROC across the studies was approximately 0.84, it was indicating a high level of accuracy. However, the calibration scores varied. Only 18% of the studies reporting well-calibrated models, which means it requires for further improvement.

F. Trust and Acceptance Among Elderly and Disable Users

The Trust and acceptance of AVs among elderly and disabled users were key themes in 14 of the studies reviewed [19] [20] [22] [23] [24] [25] [26] [30] [31] [41] [43]. These included questionnaires, behavioral experiments, and physiological data. These studies took various methods to measure trust like surveys, behavioral experiments and physiological measurements (heart rate and skin conductance). The outcomes of the studies consistently indicated that trust in AV technology was influenced by these factors:

1) The transparency of the AV's decision-making processes.
2) The predictability of the vehicle's behavior.
3) The perceived safety of the system.

Studies that integrated user feedback mechanisms into the machine learning models showed a significant improvement in trust by allowing the AV systems to adapt to individual user preferences. However, some people concerned about the potential for reduced situational awareness due to overreliance on automation, particularly among elderly users.

TABLE I. TYPES OF MACHINE LEARNING ALGORITHMS USED

Trust and Acceptance Factor	Number of Studies
Transparency of decision-making processes	14
Predictability of vehicle behavior	14
Perceived safety of the system	14
User feedback mechanisms	10
Concerns about reduced situational awareness	8

G. External Validation Bias Assessment

External validation of the machine learning models was reported in 7 of the 31 studies. These validations were conducted using independent datasets, aims to test the generalizability of the models' performance. The studies that included external validation generally reported good performance, with AUC-ROC scores consistent with those obtained during internal validation. However, in the majority of studies had the lack of external validation seems like a limitation. Users may have their concerns about the models' applicability in real-world settings. Bias assessment was addressed in 19 studies (61%), with common sources of bias being related to data collection. Especially the underrepresentation of certain demographic groups such as elderly groups and disable groups. The studies emphasized the need for more inclusive datasets to ensure that AV systems are designed to meet the needs of all users.

H. Summary of Key Findings

The findings from this scoping review indicate that significant progress has been made in applying machine learning, human-computer interaction principles, and vision neural networks. They all aim to improve the safety and usability of autonomous vehicles for diverse populations. Groups of elderly and disabled individuals were included. Vision neural networks, like those used in Tesla's current technologies, have proven essential in tasks such as object detection, road sign recognition, and pedestrian behavior prediction. These advancements have significantly made the accuracy of AV systems improved in real-world scenarios.

There are still a few challenges remain: The problems are in the areas of data quality, model validation, and user trust. Although machine learning models (including vision neural networks) have demonstrated high accuracy in controlled environments, their real-world applicability is still limited by issues. Such as data bias and the need for external validation. The integration of these advanced machine learning techniques into AV systems holds significant promise for enhancing mobility and independence among elderly and disabled individuals.

Further research is addressing the identified gaps. The key focus of fields will be in developing models that are robust, inclusive, and capable of gaining the trust of all users. Future studies should focus on improving data collection methods, increasing transparency in AV systems, and ensuring that models, including vision neural networks, are validated across diverse populations and environments.

IV. Discussion

The integration of machine learning and vision neural networks in autonomous vehicles (AVs) represents a significant leap forward in enhancing the safety, usability, and overall functionality of these technologies for diverse populations, particularly elderly and disabled users. This discussion delves into the implications of the findings, addressing the potential benefits, existing challenges, and areas for future research and development.

A. The Promise of Machine Learning and Vision Neural Networks in AVs

Machine learning algorithms, particularly those involving vision neural networks, have shown substantial potential in improving the decision-making processes of AVs. Vision neural networks enable AVs to accurately interpret visual data from the environment, facilitating real-time object detection, road sign recognition, and navigation in complex driving conditions [44] [7]. These capabilities are especially beneficial for elderly and disabled users who may face physical or cognitive limitations that impact their driving abilities [45] [46].

The findings of this review highlight the importance of these technologies in creating AV systems in both functional and inclusive. By addressing the specific needs of elderly users, such as reducing cognitive load and enhancing the predictability of vehicle behavior, machine learning models can significantly improve user trust and acceptance of AV technologies [47] [48].

B. Challenges in Data Quality and Model Validation

Despite the advancements, the review also underscores several challenges that need to be addressed to fully realize the potential of machine learning and vision neural networks in AVs. One of the primary concerns is data quality. Many studies pointed out the issues related to the underrepresentation in the datasets used to train these models. Especially in the certain demographic groups such as minorities or individuals with specific disabilities [49] [50]. This underrepresentation can lead to biased algorithms that may not perform as effectively across diverse populations [51].

Moreover, less external validation in many studies raises concerns about the generalizability of these models. The internal validation results are promising, but the models need to be tested on independent datasets. This is to ensure they can operate reliably in varied real-world conditions [45] [52]. Without this external validation, there is a risk that the models could fail when exposed to scenarios not covered in the original training data.

C. Trust and Acceptance Among Elderly Users

Trust plays a crucial role in the acceptance of autonomous vehicles (AVs), particularly among elderly users. Several studies highlight that trust, alongside transparency and perceived safety, significantly impacts how older adults interact with AV technologies [53] [54]. Machine learning models that incorporate user feedback have shown promise in enhancing trust, allowing AV systems to adapt to the specific comfort levels and preferences of elderly users [55].

However, concerns about overreliance on automation remain, with many highlighting the potential reduction in situational awareness. Maintaining a balance between offering assistance and ensuring users remain engaged is critical to preserving both safety and trust [56]. Trust-building in AV technologies requires transparency and real-time communication to help elderly passengers calibrate their trust levels accurately [54].

D. Future Research Directions

Further research is needed to optimize machine learning and vision neural networks in AVs, especially for elderly and disabled users:

1) Personalization and adaptability: Future work should focus on creating adaptable systems that cater to individual preferences, thus reducing cognitive load for elderly users [57].

2) Inclusive data collection: Researchers must ensure datasets used for training AV models include diverse user groups to mitigate biases and improve model generalizability across various populations [58].

3) Ethical and regulatory frameworks: As AV technologies evolve, regulatory frameworks must address ethical concerns, such as data privacy and algorithmic bias, to ensure AV systems remain fair and accessible [59].

4) Longitudinal studies on trust: Conducting long-term studies to track how elderly users interact with AVs can provide valuable insights into trust dynamics and safety improvements [60].

E. Implications for Industry and Policymakers

Industry stakeholders must prioritize the design of AVs that are accessible and user-friendly for elderly and disabled individuals. Investment in machine learning models that adapt to diverse user needs will be essential for broad adoption of AV technologies [61]. Additionally, policymakers should focus on creating regulatory frameworks that support the ethical development of AVs. The reason is ensuring the benefits of these technologies are equitably distributed [59].

V. Conclusion

This scoping review highlights the significant role that machine learning and vision neural networks are playing in advancing autonomous vehicle (AV) technology, particularly in addressing the needs of the aging population. These technologies make AVs to process vast amounts of data in real-time. Therefore, it's allowing for safer and more efficient driving experiences. The application of vision neural networks has improved the ability of AVs to recognize and react to objects, traffic signals, and dynamic environments in

particular, making them especially useful for elderly users who may face cognitive and physical limitations. By incorporating adaptive interfaces and assistive features, these technologies can enhance user trust, acceptance, and overall mobility, helping elderly individuals maintain their independence longer.

However, these advancements haven't achieved everything. There are still several critical challenges which need to be addressed to fully optimize AV systems for elderly populations. This scoping review revealed concerns about data quality. The underrepresentation of older adults and individuals with disabilities in datasets used for training machine learning models especially. This underrepresentation could lead to biases in the system, which may result in reduced safety and reliability for these specific user groups. Additionally, there is a limitation of external validation for many models after reviewed the articles. The raising questions about their generalizability and performance in diverse, real-world conditions.

Trust remains a key issue for the adoption of AV technology by elderly users. As AV systems become more automated, the potential for overreliance and reduced situational awareness poses a risk, especially for older adults who may already have diminished cognitive abilities. Ensuring that AV systems provide transparent and reliable feedback and allowing users to maintain a level of control when needed, is critical for building trust. Future research should prioritize developing machine learning algorithms/vision neural networks technology that can adapt to the unique cognitive and physical requirements of elderly users. The ethical concerns such as data privacy and inclusivity will also be considered.

It is essential to improve the inclusiveness of data collection efforts as the solution to resolve these challenges. The machine learning models are externally validated across diverse populations at the same time. Regulatory frameworks must evolve to address the ethical considerations surrounding AV technologies. This will include fairness, algorithmic bias, and transparency in decision-making processes. As It's particularly important as AV technology moves towards broader implementation, where ensuring accessibility for all user demographics is vital.

At present, there is a lack of interactive systems, machine learning, and visual neural network functions designed specifically for the elderly and disabled. The integration of personalized AV systems that cater specifically to the needs of elderly (as well as disabled individuals) will be essential in the coming decade to ensuring the widespread acceptance and success of AV technology. By focusing on user-centric designs which the area of HCI will exert power and continuing to improve the robustness of machine learning models. As a result, AVs with these new functions have the potential to improve the mobility, independence and quality of life for elderly populations. Policymakers, researchers and industry stakeholders must collaborate together. Their collaboration is the key of inclusive, reliable, and capable in autonomous vehicle technology. Ensuring it is reliable for the full spectrum of users' needs.

REFERENCES

[1] Parekh, D., Poddar, N., Rajpurkar, A., Chahal, M., Kumar, N., Joshi, G. P., & Cho, W. (2022). A review on Autonomous Vehicles: Progress, methods and challenges. *Electronics*, *11*(14), 2162. https://doi.org/10.3390/electronics11142162

[2] Garikapati, D., & Shetiya, S. S. (2024). Autonomous vehicles: Evolution of artificial intelligence and the current industry landscape. *Big Data and Cognitive Computing*, *8*(4), 42. https://doi.org/10.3390/bdcc8040042

[3] Shah, S. A. A., Fernando, X., & Kashef, R. (2024). A survey on artificial-intelligence-based Internet of Vehicles utilizing unmanned aerial vehicles. *Drones*, *8*(8), 353. https://doi.org/10.3390/drones8080353

[4] Garikapati, D., & Shetiya, S. S. (2024). *Autonomous vehicles: Evolution of artificial intelligence and learning algorithms*. arXiv.org. https://arxiv.org/abs/2402.17690

[5] Grigorescu, S., Trasnea, B., Cocias, T., & Macesanu, G. (2019). A survey of Deep Learning techniques for autonomous driving. *Journal of Field Robotics*, *37*(3), 362–386. https://doi.org/10.1002/rob.21918

[6] Bharilya, V., & Kumar, N. (2024). Machine learning for autonomous vehicle's trajectory prediction: A comprehensive survey, challenges, and future research directions. *Vehicular Communications*, 100733. https://doi.org/10.1016/j.vehcom.2024.100733

[7] Rani, A. R., Anusha, Y., Cherishama, S. K., & Laxmi, S. V. (2024). Traffic sign detection and recognition using Deep Learning-based approach with haze removal for Autonomous Vehicle Navigation. *E-Prime - Advances in Electrical Engineering, Electronics and Energy*, *7*, 100442. https://doi.org/10.1016/j.prime.2024.100442

[8] Ramey, J. (26AD). *Tesla Bets on AI in Latest FSD Update*. Graph and Recurrent Neural Network-based Vehicle Trajectory Prediction For Highway Driving. https://www.autoweek.com/news/a46535912/tesla-fsd-ai-neural-networks-update/

[9] Australian Institute of Health and Welfare. (2024). Older Australians. Retrieved from https://www.aihw.gov.au/reports/older-people/older-australians

[10] Faber, K., & van Lierop, D. (2020). How will older adults use automated vehicles? Assessing the role of AVs in overcoming perceived mobility barriers. *Transportation Research Part A: Policy and Practice*, *133*, 353–363. https://doi.org/10.1016/j.tra.2020.01.022

[11] Asha, A. Z., & Sharlin, E. (2023). Designing inclusive interaction with autonomous vehicles for older passengers. *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. https://doi.org/10.1145/3563657.3596045

[12] National Institute on Aging. (2022, December 20). Safe driving for older adults. https://www.nia.nih.gov/health/safety/safe-driving-older-adults

[13] Alabyad, N., Hany, Z., Mostafa, A., Eldaby, R., Tagen, I. A., & Mehanna, A. (2024). From vision to precision: The dynamic transformation of object detection in Autonomous Systems. *2024 6th International Conference on Computing and Informatics (ICCI)*. https://doi.org/10.1109/icci61671.2024.10485026

[14] Alozi, A. R., & Hussein, M. (2024). Enhancing autonomous vehicle hyperawareness in busy Traffic Environments: A machine learning approach. *Accident Analysis & Prevention*, *198*, 107458. https://doi.org/10.1016/j.aap.2024.107458

[15] Barnwal, R., Srivastava, R., R, V., & S, K. (2022). Advanced driver assistance system for autonomous vehicles using Deep Neural Network. *2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC)*. https://doi.org/10.1109/r10-htc54060.2022.9929654

[16] Bert, N., Zare, M., Larique, M., & Sagot, J. C. (2024). Design and evaluation of a virtual reality application to enhance the acceptability of autonomous vehicles for disabled people. *Learning and Analytics in Intelligent Systems*, 69–84. https://doi.org/10.1007/978-3-031-53957-2_4

[17] Borch, C., & Hee Min, B. (2022). Toward a sociology of machine learning explainability: Human–machine interaction in deep neural network-based automated trading. *Big Data & Society*, *9*(2), 205395172211113. https://doi.org/10.1177/20539517221111361

[18] Chang, J. S.-K., Chen, P.-C., Ma, H.-T., Li, S.-E., Du, W.-T., Chang, L.-H., Wang, K.-Y., Lin, C.-J., Chieh, H.-F., & Weng, C.-H. (2024).

Designing autonomous vehicle interactions for a super-aged society: A formative study. *Lecture Notes in Computer Science*, 151–167. https://doi.org/10.1007/978-3-031-61546-7_10

[19] Chen, Z. (2024). An elderly-oriented design of HMI in autonomous driving cars based on rough set theory and backpropagation neural network. *IEEE Access*, *12*, 26800–26818. https://doi.org/10.1109/access.2024.3366548

[20] Classen, S., Mason, J., Hwangbo, S. W., Wersal, J., Rogers, J., & Sisiopiku, V. (2021). Older drivers' experience with Automated Vehicle Technology. *Journal of Transport & Health*, *22*, 101107. https://doi.org/10.1016/j.jth.2021.101107

[21] Deng, Y., Xu, K., Hu, Y., Cui, Y., Xiang, G., & Pan, Z. (2022). Learning effectively from intervention for visual-based autonomous driving. *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. https://doi.org/10.1109/itsc55140.2022.9922175

[22] Eimontaite, I., Voinescu, A., Alford, C., Caleb-Solly, P., & Morgan, P. (2019). The impact of different human-machine interface feedback modalities on older participants' user experience of CAVS in a simulator environment. *Advances in Intelligent Systems and Computing*, 120–132. https://doi.org/10.1007/978-3-030-20503-4_11

[23] Gluck, A., Boateng, K., Huff Jr., E. W., & Brinkley, J. (2020). Putting older adults in the driver seat: Using user enactment to explore the design of a shared autonomous vehicle. *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. https://doi.org/10.1145/3409120.3410645

[24] Haghzare, S., Stasiulis, E., Delfi, G., Mohamud, H., Rapoport, M. J., Naglie, G., Mihailidis, A., & Campos, J. L. (2022). Automated vehicles for people with dementia: A "tremendous potential" that "has ways to go"—reports of a qualitative study. *The Gerontologist*, *63*(1), 140–154. https://doi.org/10.1093/geront/gnac115

[25] Huang, G., & Pitts, B. (2020). Age-related differences in takeover request modality preferences and attention allocation during semi-autonomous driving. *Lecture Notes in Computer Science*, 135–146. https://doi.org/10.1007/978-3-030-50252-2_11

[26] Huang, G., Hung, Y.-H., Proctor, R. W., & Pitts, B. J. (2022). Age is more than just a number: The relationship among age, non-chronological age factors, self-perceived driving abilities, and autonomous vehicle acceptance. *Accident Analysis & Prevention*, *178*, 106850. https://doi.org/10.1016/j.aap.2022.106850

[27] Isbel, S., Mulhall, S., & Gibson, D. (2022). Using Automated Vehicle Technologies with older adults: A mixed-methods study. *OTJR: Occupational Therapy Journal of Research*, *42*(3), 189–198. https://doi.org/10.1177/15394492221082493

[28] Lee, D.-H. (2024). Efficient perception, planning, and control algorithm for vision-based Automated Vehicles. *Applied Intelligence*, *54*(17–18), 8278–8295. https://doi.org/10.1007/s10489-024-05610-y

[29] Lee, H., & Samuel, S. (2024). Classification of user preference for self-driving mode and behaviors of Autonomous Vehicle. *IEEE Transactions on Intelligent Vehicles*, 1–12. https://doi.org/10.1109/tiv.2024.3385789

[30] Park, J., Zahabi, M., Blanchard, S., Zheng, X., Ory, M., & Benden, M. (2023). A novel autonomous vehicle interface for older adults with cognitive impairment. *Applied Ergonomics*, *113*, 104080. https://doi.org/10.1016/j.apergo.2023.104080

[31] Raats, K., Fors, V., & Pink, S. (2020). Trusting autonomous vehicles: An interdisciplinary approach. *Transportation Research Interdisciplinary Perspectives*, *7*, 100201. https://doi.org/10.1016/j.trip.2020.100201

[32] Reyes-Muñoz, A., & Guerrero-Ibáñez, J. (2022). Vulnerable road users and connected Autonomous Vehicles Interaction: A Survey. *Sensors*, *22*(12), 4614. https://doi.org/10.3390/s22124614

[33] Routray, S. K. (2024). Visualization and visual analytics in autonomous driving. *IEEE Computer Graphics and Applications*, *44*(3), 43–53. https://doi.org/10.1109/mcg.2024.3381450

[34] Sohail, M., Khan, A. U., Sandhu, M., Shoukat, I. A., Jafri, M., & Shin, H. (2023). Radar sensor based machine learning approach for precise vehicle position estimation. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-40961-5

[35] Sun, H., Ge, Y., & Qu, W. (2024). Greater prosociality toward other human drivers than autonomous vehicles: Human drivers' discriminatory behavior in mixed traffic. *Accident Analysis & Prevention*, *203*, 107623. https://doi.org/10.1016/j.aap.2024.107623

[36] Sun, X., Cao, S., & Tang, P. (2021). Shaping driver-vehicle interaction in autonomous vehicles: How the new in-vehicle systems match the human needs. *Applied Ergonomics*, *90*, 103238. https://doi.org/10.1016/j.apergo.2020.103238

[37] Tabone, W., de Winter, J., Ackermann, C., Bärgman, J., Baumann, M., Deb, S., Emmenegger, C., Habibovic, A., Hagenzieker, M., Hancock, P. A., Happee, R., Krems, J., Lee, J. D., Martens, M., Merat, N., Norman, D., Sheridan, T. B., & Stanton, N. A. (2021). Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. *Transportation Research Interdisciplinary Perspectives*, *9*, 100293. https://doi.org/10.1016/j.trip.2020.100293

[38] Turabian, M., Van Benthem, K., & Herdman, C. M. (2021). Electroencephalography shows effects of age in response to oddball auditory signals: Implications for semi-autonomous vehicle alerting systems for older drivers. *Lecture Notes in Computer Science*, 549–562. https://doi.org/10.1007/978-3-030-78358-7_38

[39] Wong, G. S., Goh, K. O., Tee, C., & Md. Sabri, A. Q. (2023). Review of vision-based deep learning parking slot detection on surround view images. *Sensors*, *23*(15), 6869. https://doi.org/10.3390/s23156869

[40] Xiao, Y. (2022). Application of machine learning in ethical design of autonomous driving Crash Algorithms. *Computational Intelligence and Neuroscience*, *2022*, 1–10. https://doi.org/10.1155/2022/2938011

[41] Zandieh, R., & Acheampong, R. A. (2021). Mobility and healthy ageing in the city: Exploring opportunities and challenges of autonomous vehicles for older adults' outdoor mobility. *Cities*, *112*, 103135. https://doi.org/10.1016/j.cities.2021.103135

[42] Zhang, Q., Zhang, T., & Ma, L. (2023). Human acceptance of autonomous vehicles: Research status and prospects. *International Journal of Industrial Ergonomics*, *95*, 103458. https://doi.org/10.1016/j.ergon.2023.103458

[43] Zhang, Y., Guan, J., D'Ambrosio, L. A., Miller, J., Lee, C., Zhang, K., & Coughlin, J. F. (2024). Oldest old's travel mode choice and New Mobility Technology Acceptance: Case in America and China. *Frontiers in Public Health*, *12*. https://doi.org/10.3389/fpubh.2024.1344854

[44] Kanagaraj, N., Hicks, D., Goyal, A., Tiwari, S., & Singh, G. (2021). Deep learning using computer vision in self driving cars for Lane and traffic sign detection. *International Journal of System Assurance Engineering and Management*, *12*(6), 1011–1025. https://doi.org/10.1007/s13198-021-01127-6

[45] Islam, S., Tanvir, M. S., Habib, Md. R., Shawmee, T. T., Ahmed, M. A., Ferdous, T., Arefin, Md. R., & Alam, S. (2022). Autonomous Driving Vehicle System using LIDAR SENSOR. *Lecture Notes on Data Engineering and Communications Technologies*, 345–358. https://doi.org/10.1007/978-981-16-7610-9_25

[46] Hu, X., Chen, L., Tang, B., Cao, D., & He, H. (2018). Dynamic path planning for autonomous driving on various roads with avoidance of static and moving obstacles. *Mechanical Systems and Signal Processing*, *100*, 482–500. https://doi.org/10.1016/j.ymssp.2017.07.019

[47] Linok, S. A., & Yudin, D. A. (2023). Influence of neural network receptive field on monocular depth and ego-motion estimation. *Optical Memory and Neural Networks*, *32*(S2). https://doi.org/10.3103/s1060992x23060103

[48] Li, D., & Gao, H. (2018). A hardware platform framework for an intelligent vehicle based on a driving brain. *Engineering*, *4*(4), 464–470. https://doi.org/10.1016/j.eng.2018.07.015

[49] Minaee, S., Boykov, Y. Y., Porikli, F., Plaza, A. J., Kehtarnavaz, N., & Terzopoulos, D. (2021). Image segmentation using Deep Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1. https://doi.org/10.1109/tpami.2021.3059968

[50] Hernández, Z. A., Álvarez, F., Alonso, M., & Sañudo, L. (2018). Analysis of the test criteria for vehicle containment systems in the standard EN 1317 regarding the number of vehicles in use. *Transportation Research Procedia*, *33*, 315–322. https://doi.org/10.1016/j.trpro.2018.10.108

[51] Ignatious, H. A., Sayed, H.-E.-, & Khan, M. (2022). An overview of sensors in Autonomous Vehicles. *Procedia Computer Science*, *198*, 736–741. https://doi.org/10.1016/j.procs.2021.12.315

[52] Masadeh, A., Wang, Z., & Kamal, A. E. (2018). Reinforcement learning exploration algorithms for energy harvesting communications systems. *2018 IEEE International Conference on Communications (ICC)*. https://doi.org/10.1109/icc.2018.8422710

[53] Robinson-Tay, K., & Peng, W. (2024). The role of knowledge and trust in developing risk perceptions of autonomous vehicles: a moderated mediation model. Journal of Risk Research, 1–16. https://doi.org/10.1080/13669877.2024.2360923

[54] World Economic Forum. (2024). Driving trust: Paving the road for autonomous vehicles. World Economic Forum. Retrieved from https://www.weforum.org

[55] ScienceDaily. (2024, July 9). Trust, more than knowledge, critical for acceptance of fully autonomous vehicles. ScienceDaily. Retrieved from https://www.sciencedaily.com/releases/2024/07/240709121702.html

[56] Bahrozyan, A. H. (2024). Prioritizing safety and transparency for autonomous vehicles. World Economic Forum. Retrieved from https://www.weforum.org

[57] Dong, D., Ye, H., Luo, W., Wen, J., & Huang, D. (2024). Collision avoidance path planning and tracking control for autonomous vehicles based on Model Predictive Control. *Sensors*, *24*(16), 5211. https://doi.org/10.3390/s24165211

[58] Parekh, D., Poddar, N., Rajpurkar, A., Chahal, M., Kumar, N., Joshi, G. P., & Cho, W. (2022). A review on Autonomous Vehicles: Progress, methods and challenges. *Electronics*, *11*(14), 2162. https://doi.org/10.3390/electronics11142162

[59] Lim, X. R., Lee, C. P., Lim, K. M., Ong, T. S., & Alqahtani, A. (2023). Recent advances in traffic sign recognition. Sensors, 23(10), 4674. https://doi.org/10.3390/s23104674

[60] Karle, P., Fent, F., Huch, S., Sauerbeck, F., & Lienkamp, M. (2023). Multi-modal sensor fusion and object tracking for autonomous racing. *IEEE Transactions on Intelligent Vehicles*, *8*(7), 3871–3883. https://doi.org/10.1109/tiv.2023.3271624

[61] Khan, M. A., El Sayed, H., & Malik, S. (2022). A journey towards fully autonomous driving-fueled by a smart communication system. Vehicular Communications, 36, 100476. https://doi.org/10.1016/j.vehcom.2022.100476

Machine Learning for Predicting Intradialytic Hypotension: A Survey Review

Saeed Alqahtani[1], Suhuai Luo[2], Mashhour Alanazi[3],
Kamran Shaukat[4], Mohammed G Alsubaie[5], Mohammad Amer[6]
School of Information and Physical Science, University of Newcastle, Newcastle, Australia[1, 2, 4, 5]
Prince Sultan Military College of Health Sciences, Dammam, Saudi Arabia[1]
Hemodialysis Department, King Abdulaziz City in National Guard, Riyadh, Saudi Arabia[3]
Centre for Artificial Intelligence Research and Optimisation-Design and Creative Technology Vertical,
Torrens University Australia, Ultimo, NSW 2007, Australia[4]
King Salman Armed Forces Hospital, Tabuk, Saudi Arabia[6]

***Abstract*—Intradialytic hypotension (IDH) is a common complication in patients undergoing maintenance hemodialysis and is associated with an increased risk of cardiovascular and all-cause mortality. Machine learning (ML) and deep learning (DL) techniques transform healthcare by enabling accurate disease diagnosis, personalised treatment plans, and clinical decision support. However, challenges like data quality, privacy, and interpretability must be addressed for responsible adoption. This survey review aims to summarise and analyse relevant articles on applying machine learning models for predicting IDH. Among these models, deep learning, a subfield of machine learning, stands out because it can improve the overall performance of health care, particularly in diagnostic imaging and pathologic processes and in the synthetic judgment of big data flow. The insights gained from this survey review will assist researchers and practitioners in selecting appropriate machine-learning models and implementing preemptive measures to prevent IDH in dialysis patients.**

***Keywords*—Hemodialysis; machine learning; deep learning; artificial intelligence; intradialytic hypotension; electrocardiogram; light gradient boosting machine; deep neural network; recurrent neural network**

I. Introduction

Intradialytic hypotension (IDH) is a common complication in patients undergoing maintenance hemodialysis and is associated with increased cardiovascular and all-cause mortality [1]. Intradialytic hypotension (IDH) is a frequent and serious complication of hemodialysis, characterized by a significant drop in blood pressure during dialysis sessions. Detection of IDH primarily involves continuous or frequent monitoring of blood pressure, with a drop in systolic blood pressure of ≥20 mmHg or in mean arterial pressure of ≥10 mmHg indicative of the condition [2] [3]. Advanced techniques, such as machine learning models, have been developed to predict IDH in real time by analyzing electronic health records and intradialytic data, potentially alerting clinicians 15-75 minutes before an IDH event [2]. Risk factors for IDH include patient-related factors such as age, comorbidities (e.g., diabetes, heart failure), and autonomic dysfunction, as well as dialysis-related factors like high ultrafiltration rates and incorrect target weight assessment.

Traditional solutions for managing IDH during dialysis include placing the patient in the Trendelenburg position to increase venous return, reducing or stopping ultrafiltration, and administering normal saline to restore intravascular volume [4]. Optimizing the dialysis prescription is also crucial, which can involve using cooler dialysate temperatures, adjusting sodium concentration in the dialysate, and tailoring dialysis plans based on individual patient risk factors [1][3]. Preventive strategies during the interdialytic period, such as managing interdialytic weight gain and adjusting medications, are also important [2]. Advanced techniques like blood volume monitoring and intradialytic exercise can help manage IDH effectively [4][5].

Accurate prediction of IDH is crucial for effective management and prevention of this condition. Machine learning models have shown promise in predicting IDH using various predictors and performance metrics. However, consolidating the knowledge available from the literature is necessary to gain a comprehensive understanding of the use of machine learning models for IDH prediction.

This survey review summarizes and analyses relevant articles on applying machine learning models for predicting IDH. By examining these articles, common trends can be identified, the performance of different machine learning models can be evaluated, and recommendations for future research can be provided. The insights gained from this survey review will assist researchers and practitioners in selecting appropriate machine-learning models and implementing preemptive measures to prevent IDH in dialysis patients.

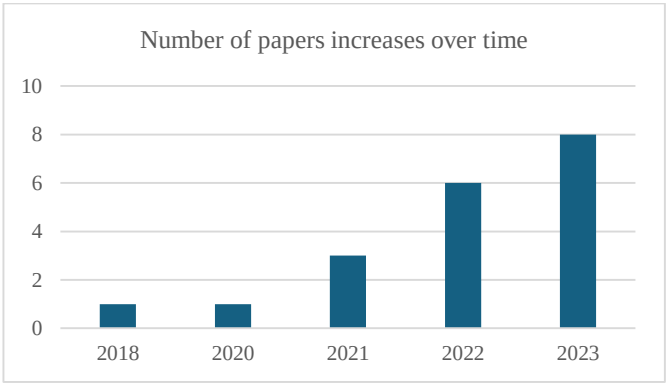


Fig. 1. The prevalence of papers using DL on IDH detection each year (Scopus).

II. Machine Learning and Intradialytic Hypotension+

Intradialytic hypotension (IDH) is one of the most frequent complications in patients requiring maintenance hemodialysis. IDH is associated with an increased risk of cardiovascular and all-cause mortality. The definition of IDH varies among studies, while the prevalence of IDH ranges up to 40% [4].

Artificial intelligence (AI) denotes computer systems capable of executing intricate tasks that traditionally were within the domain of humans, like reasoning, decision-making, or problem-solving. In contemporary usage, "AI" encompasses diverse technologies that underpin numerous services and products integral to our daily [5].

Machine learning (ML) is a branch of artificial intelligence (AI) that empowers machines to autonomously learn from data and previous experiences, enabling them to recognise patterns and formulate predictions with minimal human intervention [6].

Deep learning, a subset of machine learning, employs multi-layered neural networks known as deep neural networks to mimic the intricate decision-making processes of the human brain. This technology underpins much of the artificial intelligence (AI) we encounter daily.

Machine learning (ML) and deep learning (DL) techniques transform healthcare by enabling accurate disease diagnosis, personalised treatment plans, and clinical decision support. ML algorithms analyse patient data to predict diseases and tailor treatments. DL excels at medical image analysis, natural language processing of clinical data, genomics, and robotic surgery. While promising, challenges like data quality, privacy, and interpretability must be addressed for responsible adoption.

A deep learning architecture can be succinctly described as an Artificial Neural Network (ANN) containing two or more hidden layers to refine prediction accuracy [7]. Unlike traditional neural networks, deep learning utilises numerous hidden layers. Within a typical Deep Neural Network (DNN), input values, adjusted for weight and bias, pass through non-linear activation functions like ReLu and SoftMax to produce an output [8]. Consequently, the training objective for a DNN is to optimise network weights to minimise the loss function [9].

Traditional machine learning approaches necessitate several sequential stages, including preprocessing, feature extraction, careful feature selection, learning, and classification, to achieve classification tasks. Notably, the efficacy of machine learning techniques heavily relies on feature selection, as biased selection may yield incorrect class distinctions. Conversely, deep learning possesses the capacity to learn feature sets for diverse tasks autonomously, distinguishing it from conventional machine learning methods [10].

Although the risk factors involved in IDH are well known, including diabetes, cardiovascular disease, autonomic dysfunction, nutrition status, old age, anaemia, and high interdialytic weight gain, most of these risk factors are difficult to correct immediately at the hemodialysis centre. Therefore, treatments, such as temporarily stopping hemodialysis or reducing the ultrafiltration rate, are preferentially performed when IDH occurs. Early prediction of IDH allows for timely interventions by medical staff, such as adjusting ultrafiltration rates or administering fluids, potentially preventing or mitigating the hypotensive episode. This improves patient safety and quality of life. To detect IDH early, measuring blood pressure (BP) more frequently may be helpful. However, it is impossible to measure BP continuously due to the nature of the non-invasive BP measurement method. Thus, other non-invasive methods that can predict IDH in advance are needed.

The problem will appear when the patient's intradialytic hypotension occurs suddenly during a dialysis session. As a result, Intradialytic hypotension may reduce the efficacy of the dialysis procedure and contribute to the excessive morbidity and mortality that is associated with hemodialysis.

Artificial intelligence models have changed the paradigm of clinical decision-making from diagnosis to treatment [11]. Among these models, deep learning, a subfield of machine learning, stands out because it can improve the overall performance of health care, particularly in diagnostic imaging and pathologic processes and in the synthetic judgment of big data flow [12]. Deep learning can learn and characterise flow from a variety of data types and can thus develop a model from time-varying sequential inputs.

Fig. 2. Basic flow of the IDH system.

III. Medical Terminologies and Variables

The essential medical terms and variables frequently referenced in the 'Machine Learning for Predicting Intradialytic Hypotension' field include:

- Intradialytic Hypotension (IDH): A significant drop in blood pressure occurring during hemodialysis, typically defined as a systolic blood pressure (SBP) below 90 mmHg or a decrease of 20 mmHg or more from the baseline SBP [13].
- Mean Arterial Pressure (MAP): The average arterial pressure throughout one cardiac cycle, sometimes defined IDH as a drop of 10 mmHg or more [14].
- Nadir: The lowest point, often referring to the lowest SBP recorded during dialysis [14].
- Intradialytic SBP and MAP measurements: Real-time blood pressure readings from the dialysis machine during treatment are crucial for predicting IDH [15].

- Baseline SBP: The patient's blood pressure before starting dialysis, serving as a reference point [13].
- Ultrafiltration rate: The rate of fluid removal during dialysis can affect the risk of IDH [13].
- Demographic data: Factors like age and gender that may predict IDH risk [15].
- Clinical data: Comorbid conditions such as diabetes and cardiovascular disease increase the likelihood of IDH [15].
- Treatment data: Information on dialysis prescriptions dialysate composition [15].
- Laboratory data: Relevant blood tests, including haemoglobin levels [15].
- IDH rate: The frequency of previous IDH episodes in a patient [15].
- Area Under the Receiver Operating Characteristic (AUROC) evaluates binary classifiers by measuring their ability to distinguish positive/negative classes. It provides a scalar value summarizing the true positive/false positive trade-off across thresholds [16].
- Light Gradient Boosting Machine (LightGBM): LightGBM is an efficient, scalable, and accurate gradient boosting framework using tree-based algorithms. It employs techniques like histogram-based trees and gradient-based sampling for better performance [17].
- TabNet: TabNet is a novel deep-learning architecture designed for tabular data. It combines self-attention and sequential decision-making to capture feature interactions and learn interpretable feature importance.
- Deep Neural Network (DNN): DNNs are multi-layered neural networks capable of learning complex non-linear patterns in data for tasks like image/speech recognition but require large data and computational resources.
- Temporal Fusion Transformer (TFT): TFT is a deep learning architecture for time series forecasting, combining self-attention from transformers and temporal convolutions to capture long-range and local patterns in time series.
- Hemodialysis (HD): Hemodialysis is a life-sustaining treatment for kidney failure. It removes waste and excess fluids from the blood through a dialysis machine and artificial kidney, typically performed three times a week.
- Electrocardiogram (ECG): An ECG is a diagnostic test that measures and records the electrical activity of the heart. It is used to detect and monitor heart conditions by analysing electrical signals from electrodes.

These terms and variables, especially the real-time intradialytic blood pressure measurements, are often used as inputs for machine learning models to predict the risk of IDH during hemodialysis.

In the following paragraphs, an overview of research papers on Intradialytic hypotension will be presented, focusing on scientific papers that have utilised machine learning methods to predict Intradialytic hypertension before its occurrence. The objective is to identify, detect, analyse, and evaluate the models used for detecting Intradialytic hypertension, ultimately determining the best model to enhance patient health and improve the quality of medical care provided.

IV. Threat to Validity: Search Strategy and Databases

In this survey review, we acknowledge the importance of providing a detailed explanation of the search strategy and the databases used to address the threat to validity. By employing a rigorous approach to literature search and retrieval, we aimed to capture a wide range of relevant studies and provide valuable insights for researchers and practitioners in the field of machine learning for predicting intradialytic hypotension.

A comprehensive survey was conducted to gather information on using machine learning models for predicting intradialytic hypotension (IDH). The survey design involved identifying relevant articles published in the literature. The target population for the survey review included studies that investigated the application of machine learning models for predicting IDH in patients undergoing maintenance hemodialysis.

A. Search Strategy and Databases

The survey review employs a wide-ranging search approach to ensure that multiple databases are utilised to capture a diverse range of studies. The following databases play a crucial role in the search process:

1) PubMed: This widely recognised and extensively used database for biomedical literature contains a vast collection of articles from various disciplines, including medicine, biology, and healthcare.

2) Scopus: As a comprehensive bibliographic database, Scopus covers a wide range of scientific disciplines. It includes articles from peer-reviewed journals, conference proceedings, and patents, providing a broad scope of literature for research purposes.

3) Google scholar: Designed as a web search engine for scholarly literature, Google Scholar indexes academic publications from various sources, including journals, conference papers, and theses.

B. Search String

The search string employed in the survey review consists of relevant keywords and phrases related to machine learning, intradialytic hypotension, and associated techniques. The specific search terms are tailored to each database's syntax and search capabilities, ensuring optimal retrieval of relevant articles.

The following search terms and combinations were used:

1) "Intradialytic hypotension" AND "machine learning"
2) "Intradialytic hypotension" AND "predictive models"
3) "Machine learning" AND "hemodialysis" AND "hypotension"
4) "Predictive modelling" AND "intradialytic hypotension"
5) "Deep learning," "artificial intelligence,"
6) Specific machine learning algorithms like "decision trees," "support vector machines," and "random forest."

By utilising these databases and employing a comprehensive search string, the survey review aims to minimise the possibility of missing relevant studies and provide a comprehensive overview of the use of machine learning models for predicting intradialytic hypotension. The search strategy ensures a thorough examination of the available evidence and current trends in the field.

C. Inclusion / Exclusion Criteria

The inclusion criteria for study selection were as follows:

1) Articles focused on applying machine learning models to predict IDH in patients undergoing hemodialysis maintenance.
2) Articles published in the English language.

Articles that did not meet the inclusion criteria or were inaccessible for data extraction were excluded.

Articles were excluded if they did not meet the inclusion criteria or were inaccessible for data extraction. The articles that met the inclusion criteria were included in the analysis to provide a comprehensive overview of the use of machine learning models for predicting IDH.

After the initial search, the study selection process began. The selection was carried out in two stages: title/abstract screening and full-text screening. During the title/abstract screening stage, articles that did not meet the inclusion criteria or were irrelevant were excluded. The remaining articles underwent full-text screening, where they were assessed for eligibility based on the predefined inclusion/exclusion criteria.

D. Included Studies

Of 270 articles retrieved from PubMed, Scopus and Google Scholar, 200 were ruled out after title screening, 43 were excluded after abstract screening, 26 were selected for full-text screening, and 12 articles were deemed pertinent and included in the final analysis. These 12 articles are summarised in Table I, Machine learning for IDH.

We created two tables: Table I presented the models used in all studies of intradialytic hypotension disease. Table II presents all studies that used ML to detect IDH disease.

TABLE I. INCLUDES MODELS / METHODS OF 12 ARTICLES SHOWING MACHINE LEARNING USED IN THE SELECTED PAPERS

No	Article name	ML models used	Best model used	Models' contribution
[1]	Predicting the Appearance of Hypotension during Hemodialysis Sessions Using Machine Learning Classifiers	Decision Trees (DT) and Support Vector Machines (SVM). The prediction model achieved a success rate in prediction higher than 80%	The Decision Trees (DT) model achieved slightly better results in predicting hypotension during hemodialysis compared to the Support Vector Machines (SVM) model, with mean accuracy rates ranging from 75% to 81% for DT and 74% to 80% for SVM	The Decision Trees (DT) model achieved the goal by generating predictive models from the clinical parameters as attributes or input variables, categorising data into finite classes, and accurately predicting the occurrence of hypotension during hemodialysis sessions.
[2]	Real-time prediction of intradialytic hypotension using machine learning and cloud computing infrastructure	The ML models used to predict intradialytic hypotension are a machine learning model developed by Zhang et al. and a deep learning model developed by Lee et al.	The model developed by Lee et al. achieved the goal with an area under the receiver operating characteristic curve of 0.94characteristic curve (AUROC) serving as an indicator of predictive efficacy.	The machine learning model developed by Zhang et al. contributes to predicting intradialytic hypotension by utilising electronic health records comprising intradialytic blood pressure measurements and multiple treatment- and patient-level variables to generate alerts before an IDH event, facilitating timely interventions to prevent it.
[3]	Machine Learning Analysis of Time-Dependent Features for Predicting Adverse Events During Hemodialysis Therapy: Model Development and Validation Study	The types of algorithms mentioned are: - Bayes point machine, Boosted decision tree, SVM (Support Vector Machine), Two-class average perceptron, Deep learning, Two-class logistic regression and Decision forest.	Lin et al. developed an intelligent system that achieved the goal of predicting intradialytic hypotension in chronic hemodialysis patients.	This system contributes to predicting intradialytic hypotension by analysing data and identifying patterns that indicate a risk of hypotension during dialysis sessions, thereby potentially allowing for preemptive measures to prevent this complication.
[4]	Construction of an Early Alert System for Intradialytic Hypotension before Initiating Hemodialysis Based on Machine Learning	Random forest, gradient boosting, and logistic regression were the three best models used to predict intradialytic hypotension.	Among them, the Random Forest model achieved the goal with an AUC of 0.812 (95% CI, 0.811 – 0.813).	The Random Forest model achieved the goal of predicting intradialytic hypotension by having the highest AUC value of 0.812, which was the best among the models tested.
[5]	Machine learning-based intradialytic	Machine learning models used to predict intradialytic	Among these models, the LightGBM method was identified as the best-	LightGBM contributed to predicting IDH risk in hemodialysis patients by providing a

	hypotension prediction of patients undergoing hemodialysis: A multicenter retrospective study	hypotension (IDH) included Light Gradient Boosting Machine (LightGBM), Linear Discriminant Analysis, support vector machines, XGBoost, TabNet, and multilayer perceptron.	performing and interpretable model with a C-statistic of 0.82 in Fall30Nadir90 definitions, which was higher than those obtained using other models.	reliable risk prediction model for the next hemodialysis treatment.
[6]	Deep Learning for Intradialytic Hypotension Prediction in Hemodialysis Patients	linear regression method and a deep neural network (DNN).	Among these, the deep neural network achieved the goal of predicting IDH with the maximum performance metrics evaluation.	This indicates that the DNN model was a potential tool for the management of intradialytic hypotension by effectively predicting its occurrence during a hemodialysis session.
[7]	Deep Learning Model for Real-Time Prediction of Intradialytic Hypotension	Random forest, gradient boosting, and logistic regression were the three best models used to predict intradialytic hypotension.	Among them, the Random Forest model achieved the goal with an AUC of 0.812 (95% CI, 0.811 – 0.813) The Random Forest model was identified as the best model for predicting intradialytic hypotension with an AUC of 0.812, which was the highest among the models tested.	The Random Forest model achieved the goal of predicting intradialytic hypotension by having the highest AUC value of 0.812, which was the best among the models tested.
[8]	Machine-Learning Model to Predict the Intradialytic Hypotension Based on Clinical- Analytical Data	XGBoost model and model based on a recurrent neural network (RNN).	The XGBoost model was identified as achieving the goal of predicting IDH at the beginning of the hemodialysis (HD) session with high predictive capacity, achieving values of 0.969 and 0.945 for areas under the receiver operating characteristic (AUROC)	These models contribute to predicting IDH by identifying complex relationships among clinical-analytical data that traditional statistical models cannot, thereby supporting informed decision-making by medical staff.
[9]	Prediction of intradialytic hypotension using pre-dialysis features-a deep learning-based artificial intelligence model	The machine learning (ML) models used to predict intradialytic hypotension (IDH) include logistic regression, random forest, XGBoost, and a deep learning model based on a one-dimensional convolutional neural network (1D-CNN).	Among these, the deep learning-based artificial intelligence (AI) model achieved superior performance compared to the previously reported results.	This deep learning-based AI model contributes to predicting IDH accurately by using pre-dialysis features, thus allowing clinicians to identify vulnerable patients and adjust hemodialysis (HD) settings before HD treatment.
[10]	Deep Learning Model for Predicting Intradialytic Hypotension Without Privacy Infringement: A Retrospective Two-Center Study	The machine learning (ML) models used to predict intradialytic hypotension (IDH) include logistic regression, random forest, XGBoost, and convolutional neural networks (CNNs).	Among these, the deep learning model, specifically using convolutional neural networks, achieved the best performance in predicting IDH. It performed well only using monitoring measurements from the hemodialysis machine during the dialysis session, without requiring any personal information that could risk privacy infringement.	This contributes to predicting IDH by allowing for effective monitoring and management of the condition in real-time, potentially reducing the risk associated with IDH.
[11]	Real-time dual prediction of intradialytic hypotension and hypertension using an explainable deep learning model	The machine learning (ML) models used to predict intradialytic hypotension include the Temporal Fusion Transformer (TFT), Recurrent Neural Network (RNN), and Light Gradient Boosting Machine (LightGBM).	Among these, the Temporal Fusion Transformer (TFT)-based model was specifically developed and utilised to predict both intradialytic hypotension and h ypertension simultaneously, achieving the goal of real-time dual prediction.	This TFT-based model contributes to predicting intradialytic hypotension by leveraging a sequence-to-sequence-based attention network, which allows for the simultaneous prediction of intradialytic hypotension and hypertension in real-time, utilising data from electronic health records of hemodialysis sessions.
[12]	Deep Learning on Electrocardiograms for Prediction of In-hospital Intradialytic Hypotension in Patients with ESKD	Densenet-201 model pretrained on a publicly available data set of natural images.	Densenet-201 model pretrained on a publicly available data set of natural images.	In summary, using deep learning on ECG waveform data, we were able to predict with moderate accuracy which patients would develop IDH at their next HD treatment.

Table I presents all models/methods used to diagnose IDH. We selected the best models after the presentation of all used models in Table I. Then we highlighted the models/methods contributions for IDH patients.

V. State-of-the-Art Approaches

A. Overview of ML Models for IDH Patients

Machine Learning (ML) is a branch of Artificial Intelligence that enables computers to learn and improve from experience without being explicitly programmed. It involves developing algorithms and statistical models that allow systems to perform specific tasks effectively by leveraging patterns and insights from data. ML has revolutionized various fields, including computer vision, natural language processing, and predictive analytics. Table II presents studies that employed machine learning methodologies tailored specifically to predict and manage a distinct medical condition—intradialytic hypotension (IDH). These methodologies were predominantly geared

towards multifaceted objectives encompassing the prediction of hypotension onset during hemodialysis sessions, anticipation of IDH occurrences in in-centre hemodialysis patients with a lead time of 15–75 minutes, risk prediction for patients undergoing hemodialysis, unbiased prediction of intradialytic adverse events, and early identification of patients prone to IDH development at the onset of hemodialysis sessions. Notably, the breadth of purposes addressed by most studies included developing or utilising models and algorithms to enhance clinical decision-making and patient outcomes in managing IDH. Moreover, while most studies focused on algorithmic advancements or predictive modelling, one notable paper stands out for its novel approach: employing machine learning to construct a new system to manage intradialytic hypotension.

TABLE II. OVERVIEW OF ALL STUDIES THAT USED ML TO DETECT IDH DISEASE

No	Article name	Objective	Method	Conclusions
[1]	Predicting the Appearance of Hypotension during Hemodialysis Sessions Using Machine Learning Classifiers	This article explores the prediction of hypotension during dialysis sessions by utilizing predictive models trained on a comprehensive dialysis database, which includes clinical data from 98,015 sessions involving 758 patients.	The forecasting algorithm considers 22 clinical factors assessed five times throughout the session, along with the patient's gender and age. It was trained using machine learning classifiers and achieves a prediction accuracy of over 80%.	Using machine learning techniques, the researchers pinpointed lightgbm as an effective and understandable model. They also identified key variables indicating high-risk factors for IDH incidents in HD patients. The IDH-A and IDH-B models can complement each other, enhancing risk prediction and enabling timely intervention across various clinical settings.
[2]	Real-time prediction of intradialytic hypotension using machine learning and cloud computing infrastructure.	This article aims to develop a machine learning model to predict IDH in in-centre hemodialysis patients 15–75 minutes in advance.	They constructed a machine learning framework to anticipate IDH occurrences in in-centre hemodialysis patients, predicting events 15–75 minutes in advance based on systolic blood pressure (SBP) falling below 90 minutes. Data from electronic health records encompassing demographic, clinical, treatment-related, and laboratory information were merged with real-time intradialytic machine data transmitted to the cloud. Dialysis sessions were randomly divided into training (80%) and testing (20%) subsets for model development, with the area under the receiver operating characteristic curve (AUROC) serving as an indicator of predictive efficacy.	The feasibility of real-time IDH prediction during hemodialysis sessions demonstrates clinically relevant predictive capability. Prospective studies are necessary to assess the extent to which this predictive insight aids in deploying preventive measures promptly, potentially leading to reduced IDH rates and improved patient outcomes.
[3]	Construction of an Early Alert System for Intradialytic Hypotension before Initiating Hemodialysis Based on Machine Learning	This article aims to explore the risk factors for IDH and use artificial intelligence to establish an early alert system before hemodialysis sessions to identify patients at high risk of IDH.	Researchers employed four interpolation techniques, three feature selection approaches, and 18 machine learning algorithms to build predictive models. Model performance evaluation relied primarily on the area under the receiver operating characteristic curve (AUC), while Shapley Additive Explanation was utilised to elucidate the impact of each variable on the most effective predictive model.	Our artificial intelligence alert system, integrated into dialysis software, can forecast the onset of intradialytic hypotension, facilitating timely intervention strategies.
[4]	Machine learning-based intradialytic hypotension prediction of patients undergoing hemodialysis: A multicenter retrospective study.	This article used Machine learning (ML) to develop an IDH model for risk prediction in HD patients. Six candidate machine learning methods, such as lightGBM [17], support vector machine (SVM) [18], linear discriminant analysis (LDA) [19], eXtreme Gradient Boosting (XGB) [20], TabNet [21], and multilayer perceptron (MLP) [22] were utilised for prediction models.	62,227 dialysis sessions were randomly divided into training (70%), testing (20%), and validation (10%) datasets. An IDH-A model, incorporating twenty-seven variables, was constructed to predict the risk of intradialytic hypotension (IDH) for subsequent hemodialysis (HD) treatments. Additionally, an IDH-B model, utilising ten variables from 64,870 dialysis sessions, was developed to assess the risk of IDH before each HD session. The predictor model was constructed using a Light Gradient Boosting Machine (lightgbm), Linear Discriminant Analysis, support vector machines, xgboost, tabnet, and multilayer perceptron algorithms.	Using ML, we identified the lightgbm method as the good-performing and interpretable model. We identified the top variables as the high-risk factors for IDH incidents in HD patients. IDH-A and IDH-B models can usefully complement each other for risk prediction and further facilitate timely intervention in different clinical settings.

[5]	Machine Learning Analysis of Time-Dependent Features for Predicting Adverse Events During Hemodialysis Therapy: Model Development and Validation Study.	This article seeks to create machine learning algorithms that can impartially predict adverse events occurring during dialysis.	Over a three-month period, data were collected from all patients receiving maintenance hemodialysis at a tertiary care referral center, encompassing dialysis and physiological time-series records. Hemodialysis machines automatically captured dialysis data, while medical staff documented physiological information and recorded intradialytic adverse events based on patient reports. Features derived from the time-series data using linear and differential analyses were employed in machine learning to predict adverse events during hemodialysis.	Our findings reveal that algorithms combining linear and differential analyses with two-class classification machine learning can accurately predict intradialytic adverse events in real-time. Utilizing this approach with local cloud computing and personalized hemodialysis data for real-time optimization could enable clinicians to take proactive measures swiftly.
[6]	Machine-Learning Model to Predict the Intradialytic Hypotension Based on Clinical-Analytical Data	This article details the development, evaluation, and identification of a machine learning-based model that can predict, at the outset of a hemodialysis session, whether a patient will experience intradialytic hypotension as the session unfolds.. In this research study, different ML models (Logistic Regression, Random Forest, Multi-layer Perceptron and eXtreme Gradient Boosting [XGBoost])	They utilized both hold-out and cross-validation methods to develop their machine learning models. To evaluate the models' performance, they employed metrics such as F1-score, Matthews Correlation Coefficient, area under the receiver operating characteristic (AUROC) curve, and area under the precision-recall curve (AUPRC). They carefully selected a subset of variables from clinical records and blood analyses that were essential for predicting intradialytic hypotension (IDH).	Among the ML models tested, the xgboost model demonstrated superior performance, achieving highly reliable results with accuracy, precision, recall, F1-score, and MCC values of 0.92, 0.90, 0.82, 0.86 and 0.81, respectively.

Gómez-Pulido et al. (2021) conducted a study titled "Predicting the Appearance of Hypotension during Hemodialysis Sessions Using Machine Learning Classifiers" to discern the occurrence of hypotension during dialysis sessions [20]. Leveraging a comprehensive dialysis database encompassing 98,015 sessions of 758 patients, the researchers employed predictive models trained on 22 clinical factors evaluated five times per session alongside demographic attributes. Employing machine learning classifiers, their methodology achieved a prediction accuracy surpassing 80%. The study underscored the efficacy of lightgbm as an interpretable model and identified critical variables indicative of high-risk factors for Intradialytic Hypotension (IDH) incidents in hemodialysis patients. Furthermore, the study revealed the complementary nature of IDH-A and IDH-B models, augmenting risk prediction and facilitating timely interventions across diverse clinical settings.

In a parallel endeavour, Liu et al. (2021) developed machine learning algorithms for unbiased prediction of intradialytic adverse events. The study meticulously collected dialysis and physiological time-series records using data spanning three months from patients undergoing maintenance hemodialysis. These datasets were subjected to linear and differential analyses to extract features conducive to machine learning-based prediction of adverse events during hemodialysis. Notably, integrating linear and differential analyses with two-class classification machine learning yielded near-real-time prediction of intradialytic adverse events with high AUCs. The study suggests the potential implementation of such methodologies, augmented by local cloud computation and personalised hemodialysis data, for real-time optimisation and proactive clinical interventions.

Mendoza-Pittí et al. (2022) focused on developing, evaluating, and identifying an ML-based model capable of predicting the onset of IDH at the outset of hemodialysis sessions. Employing hold-out and cross-validation techniques, the researchers meticulously assessed model performance using metrics such as F1-score, Matthews Correlation Coefficient, and areas under the receiver operating characteristic and precision-recall curves. By carefully selecting and utilising a subset of variables from clinical records and blood analytics, the study identified the xgboost model as a superior performer, exhibiting highly reliable predictive capabilities with notable accuracy and precision metrics.

Continuing this trajectory, Gómez Zhang et al. (2023) aimed to predict IDH occurrences in in-centre hemodialysis patients 15–75 minutes in advance. Leveraging electronic health records merged with real-time intradialytic machine data. The researchers constructed a machine learning framework for anticipatory IDH prediction. Their methodology, validated using a split of training and testing subsets, demonstrated clinically relevant predictive capability. Prospective studies are warranted to gauge the translational potential of such predictive insights in clinical practice, potentially leading to reduced IDH rates and enhanced patient outcomes.

Daqing Hong 1 [21] contributed to explore risk factors for IDH and establish an early alert system using artificial intelligence. Employing various interpolation techniques, feature selection approaches, and 18 machine learning algorithms, the study evaluated model performance using the area under the receiver operating characteristic curve. The integration of artificial intelligence into dialysis software facilitated the forecasting of IDH onset, enabling timely interventions to mitigate adverse outcomes.

Dong et al. (2023) developed IDH prediction models for HD patients. Through meticulous analysis of 62,227 dialysis sessions, the researchers developed IDH-A and IDH-B models, leveraging Light Gradient Boosting Machine (lightgbm), Linear Discriminant Analysis, support vector machines, xgboost, tabnet, and multilayer perceptron algorithms. The study identified lightgbm as a performant and interpretable model, shedding light on high-risk factors for IDH incidents. The complementarity of IDH-A and IDH-B models underscores

their potential for risk prediction and timely interventions in diverse clinical settings.

B. Machine Learning Insights for IDH Patients

The inception of this investigative trajectory was marked by Gómez-Pulido et al. (2021), who utilised Decision Trees (DT) and Support Vector Machines (SVM) to forecast hypotension events, with DT models demonstrating a superior classification capability. This study underscored the criticality of algorithm selection in clinical prediction tasks. Advancing the discourse, Liu et al. (2021) employed a suite of algorithms, including LightGBM and XGBoost, with LightGBM distinguishing itself as the most reliable for IDH risk assessment. This revelation accentuates the potential of gradient-boosting methods in enhancing clinical decision support systems. Lin et al.'s exploration of various algorithms, including Bayes point machine and Two-class logistic regression, culminated in a system adept at identifying IDH risk patterns. This versatility in algorithm application indicates ML's robustness in healthcare analytics. The studies by Hong et al. (2023) and Dong et al. (2023) further corroborate the efficacy of Random Forest and LightGBM models, respectively, in IDH prediction. Notably, the Random Forest model's AUC value of 0.812 achieved by Hong et al. (2023) is emblematic of the model's predictive precision. Zhang et al. (2023) concluded the ML narrative by leveraging time-dependent features from electronic health records.

C. Overview of DL Models for IDH Patients

Deep Learning (DL) is a subset of Machine Learning that uses artificial neural networks inspired by the human brain's structure and function. It involves training multi-layered neural networks on vast amounts of data to learn hierarchical representations and patterns. DL has achieved remarkable success in areas like image recognition, speech recognition, natural language processing, and reinforcement learning.

Table III shows studies that employed deep learning methodologies tailored specifically to the prediction and management of a distinct medical condition, namely, intradialytic hypotension (IDH) as follows:

TABLE III. OVERVIEW FOR ALL STUDIES THAT USED DL TO DETECT IDH DISEASE

No	Article name	Objective	Method	Conclusions
[1]	Deep Learning Model for Real-Time Prediction of Intradialytic Hypotension	This article utilized a timestamped dataset to apply a deep learning model, specifically a recurrent neural network, for predicting the risk of intradialytic hypotension.	They gathered data from 261,647 hemodialysis sessions, comprising 1,600,531 unique timestamps reflecting time-varying vital signs. These sessions were randomly split into training (70%), validation (5%), calibration (5%), and testing (20%) sets. The performance of the recurrent neural network model was evaluated against that of multilayer perceptron, Light Gradient Boosting Machine, and logistic regression models, using metrics such as the area under the receiver operating characteristic curves, the area under the precision-recall curves, and F1 scores.	Our deep learning model allows for the real-time anticipation of intradialytic hypotension risk.
[2]	Deep Learning for Intradialytic Hypotension Prediction in Hemodialysis Patients	This article investigates clinical factors associated with intradialytic hypotension by deep learning.	In March 2018, 279 participants undergoing outpatient hemodialysis at a hospital-based center were included, accounting for 780 hemodialysis sessions. Relationships between clinical variables and intradialytic hypotension were assessed using linear regression and deep neural network methodologies. The predictive model generated through deep learning techniques holds promise as a potential tool for managing intradialytic hypotension.	This study underscores the potential of a deep learning approach in identifying the clinical factors associated with intradialytic hypotension during hemodialysis sessions.
[3]	Prediction of intradialytic hypotension using pre-dialysis features a deep learning-based artificial intelligence model	This article aims to develop a deep learning-based artificial intelligence (AI) model to predict IDH using pre-dialysis features.	Information was utilised from 2007 patients undergoing 943,220 hemodialysis (HD) sessions across seven university hospitals. The deep learning model's effectiveness was evaluated compared to three other machine learning models (logistic regression, random forest, and XGBoost).	Our artificial intelligence (AI) model demonstrates accurate prediction of intradialytic hypotension (IDH), indicating its reliability as a valuable tool for guiding hemodialysis (HD) treatment.
[4]	Deep Learning Model for Predicting Intradialytic Hypotension Without Privacy Infringement: A Retrospective Two-Center Study	Previously developed prediction models for Intradialytic hypotension (IDH) typically rely on clinical variables that may raise privacy concerns. In response, they devised an IDH prediction model using minimal variables to mitigate the risk of privacy infringement.	The models they developed utilize 30-minute data to predict an intradialytic hypotension (IDH) event within 10 minutes. They compared various machine learning and deep learning models, including logistic regression, XGBoost, and convolutional neural networks, using metrics such as the area under the receiver operating characteristic curves (AUROCs) and precision-recall curves.	In comparison with models incorporating logistic regression, random forest, and XGBoost, the deep learning model exhibited superior performance in predicting IDH. Specifically, the deep learning model achieved the highest performance metrics by solely utilizing measurements from the hemodialysis machine during the dialysis session, with AUROCs of 0.905 for Nadir90, 0.864 for Fall20, and 0.863 for Fall20/MAP10.

[5]	Real-time dual prediction of intradialytic hypotension and hypertension using an explainable deep-learning model	This article aims to develop an explainable deep learning model with a sequence-to- sequence-based attention network to simultaneously predict both events.	Data from 11,110 patients, encompassing 302,774 hemodialysis sessions, were retrieved from electronic health records across seven university hospitals. These sessions were divided randomly into training (70%), validation (10%), and test. (20%) datasets. The performance of the developed deep learning model, along with other machine learning models such as logistic regression, random forest, and XGBoost, was evaluated using metrics like AUROC and AUPRC.	Among all models, the TFT-based model achieved the highest AUROCs of 0.953 (0.952-0.954), 0.892 (0.891-0.893), and 0.889 (0.888-0.890) in predicting IDH- 1, IDH-2, and IDHTN, respectively. The AUPRCs in theTFT- based model for these outcomes were higher than the other models. The factors that contributed the most to the prediction were age and previous session, as well as systolic BP and elapsed time, which were time-varying variables.
[6]	Deep Learning on Electrocardiograms for Prediction of In-hospital Intradialytic Hypotension in Patients with ESKD	This article aims to develop a deep learning framework on ECG waveform data to predict in-hospital IDH.	To study intradialytic hypotension, they collected the starting and ending times of each patient's inpatient hemodialysis procedures and blood pressure readings. In addition, they extracted electrocardiograms (ECGs) performed 48 hours before each hemodialysis session and utilised 2D Convolutional Neural Network analysis on ECG waveform data paired with intradialytic hypotension occurrences.	In summary, using deep learning on ECG waveform data, They could predict with moderate accuracy which patients would develop IDH at their next HD treatment. This ECG-based model performed well across all predialysis SBP subgroup analyses with slight differences in AUROC.

Initially, Chen et al. (2020) contributed to this domain with their work titled "Deep Learning for Intradialytic Hypotension Prediction in Hemodialysis Patients," aiming to discern clinical factors associated with intradialytic hypotension using deep learning methodologies [22]. Their investigation involved 279 participants undergoing outpatient hemodialysis at a hospital-based centre, comprising 780 hemodialysis sessions. Relationships between clinical variables and IDH were scrutinised via linear regression and deep neural network approaches. The resultant predictive model, forged through deep learning techniques, emerged as a promising tool for managing intradialytic hypotension, thus underscoring the potential of deep learning in identifying pertinent clinical factors linked to IDH occurrences during hemodialysis sessions.

Lee et al. (2021) aiming to forecast the risk of intradialytic hypotension (IDH) utilising a timestamp-rich dataset [14]. Their investigation encompassed data from 261,647 hemodialysis sessions, capturing 1,600,531 distinct timestamps representing dynamic, vital signs. The dataset was meticulously partitioned into training (70%), validation (5%), calibration (5%), and testing (20%) subsets. Performance assessments of the recurrent neural network model were juxtaposed against those of multilayer perceptron, Light Gradient Boosting Machine, and logistic regression models, employing metrics such as the area under the receiver operating characteristic curves, an area under the precision-recall curves, and F1 scores. The findings elucidated the efficacy of deep learning in enabling real-time prediction of IDH risk.

Kim et al. (2022) developed IDH prediction models while safeguarding patient privacy [13]. Employing machine learning and deep learning models, including logistic regression, XGBoost, and convolutional neural networks, they evaluated model performance using area under the receiver operating characteristic curves (AUROCs) and precision-recall curves. Their findings indicated the superiority of the deep learning model in predicting IDH, particularly when utilising measurements solely from the hemodialysis machine during sessions.

Lee et al. (2023) leveraged data from 2007 patients undergoing 943,220 hemodialysis sessions across seven university hospitals [23]. The efficacy of the deep learning model was benchmarked against three other machine learning models—logistic regression, random forest, and XGBoost. The outcomes underscored the artificial intelligence model's aptitude for accurate IDH prediction, positioning it as a reliable tool for guiding hemodialysis treatment decisions.

Furthermore, Yun et al. (2023) incorporated data from 11,110 patients and 302,774 hemodialysis sessions, partitioned into training, validation, and test datasets. The developed deep learning model, alongside other machine learning models, demonstrated robust performance metrics, with the TFT-based model exhibiting the highest AUROCs and AUPRCs among all models tested.

Lastly, Vaid et al. (2023) developed a deep learning framework leveraging ECG waveform data to forecast in-hospital IDH [24]. Their investigation utilised ECGs performed 48 hours before each hemodialysis session, employing 2D Convolutional Neural Network analysis on ECG waveform data paired with intradialytic hypotension occurrences. The findings highlighted the model's moderate accuracy in predicting which patients would develop IDH at their next HD treatment, with consistent performance across all predialysis SBP subgroup analyses.

Most included articles (n = 26) were published from 2020 to 2023. Among the 68 articles selected for data extraction, most were published by authors from organisations based in 2020 (n = 15; 58%). The remaining articles were published by authors in 2021 (n = 4; 15%), in 2022 (n = 2; 8%) and 2023 (n = 5; 19%). The analysed studies were classified as observational.

D. Deep Learning Insights for IDH Patients

The foray into DL was pioneered by Chen et al. (2020), whose comparative analysis between linear regression and a deep neural network (DNN) favoured the latter, illustrating DL's adeptness in managing complex clinical datasets. Subsequent research by Lee et al. (2021) yielded a Random Forest model with an AUC value of 0.812, reinforcing the applicability of ensemble methods in real-time IDH prediction. Kim et al. (2022) and Lee et al. (2023) emphasised the utility of convolutional neural networks (CNNs) in IDH prediction, with the former

addressing privacy concerns. The CNN model's real-time monitoring capability represents a significant stride in patient-centric care. Yun et al. (2023) introduced an explainable deep learning model employing a Temporal Fusion Transformer (TFT) for dual IDH and hypertension prediction, showcasing the model's multifaceted predictive capacity. Vaid et al. (2023) explored the use of DL on electrocardiograms for IDH prediction in ESKD patients, contributing to the burgeoning evidence of DL's clinical relevance.

Numerous endeavours have been undertaken to enhance the prediction of Intradialytic Hypotension (IDH) by applying Machine Learning (ML) models. The following highlights several notable efforts in this domain:

Initially, Gómez-Pulido et al. (2021) embarked on predicting the onset of hypotension during hemodialysis sessions employing Machine Learning Classifiers, specifically Decision Trees (DT) and Support Vector Machines (SVM). Their investigation revealed that the DT model outperformed others in generating predictive models from clinical parameters, effectively categorising data into distinct classes and accurately anticipating the incidence of hypotension during hemodialysis sessions.

Subsequently, Liu et al. (2021) conducted a similar multicenter retrospective study employing LightGBM, Linear Discriminant Analysis, support vector machines, XGBoost, TabNet, and multilayer perceptron models. Consistently, the LightGBM method emerged as the optimal choice for predicting IDH risk in hemodialysis patients, offering a dependable risk assessment tool for guiding clinical decision-making.

In a parallel effort, Liu et al. (2021) delved into the prediction of hypotension during hemodialysis sessions utilising a diverse array of algorithms, including Bayes point machine, Boosted decision tree, SVM, Two-class average perceptron, Deep learning, Two-class logistic regression, and Decision Forest. Among these, the system developed by Lin et al. emerged as the most intelligent, effectively discerning patterns indicative of hypotension risk during dialysis sessions, thereby enabling proactive measures to avert this complication.

Furthermore, Hong et al. (2023) endeavoured to construct an Early Alert System for Intradialytic Hypotension based on Machine Learning, employing Random forest, gradient boosting, and logistic regression models. Of these, the Random forest model demonstrated superior performance, achieving the highest Area Under the Curve (AUC) value of 0.812, thereby exhibiting enhanced predictive capability for IDH.

Additionally, Dong et al. (2023) conducted a multicentre retrospective study to predict IDH risk in hemodialysis patients utilising a suite of models including Light Gradient Boosting Machine (LightGBM), Linear Discriminant Analysis, support vector machines, XGBoost, TabNet, and multilayer perceptron. Their findings underscored the efficacy of the LightGBM method in furnishing a robust risk prediction model for subsequent hemodialysis treatments.

Lastly, Zhang et al. (2023) pursued using Machine Learning Analysis of Time-Dependent Features for Predicting Adverse Events During Hemodialysis Therapy. Employing models developed by Zhang et al. and Lee et al., their study identified the deep learning model by Lee et al. as the most efficacious. This model leveraged electronic health records encompassing intradialytic blood pressure measurements and various treatment- and patient-level variables to pre-emptively generate alerts preceding IDH events, thereby facilitating timely interventions for prevention.

These endeavours collectively illuminate the potential of ML models in advancing the prediction and management of IDH, underscoring the significance of ongoing research in this critical healthcare domain.

In the pursuit of advancing the prediction of Intradialytic Hypotension (IDH) through Deep Learning (DL) models, several studies have contributed valuable insights. The following outlines notable endeavours in this area:

Chen et al. (2020) initiated investigations into the application of Deep Learning for Intradialytic Hypotension Prediction in Hemodialysis Patients. Their study employed a comparative analysis between a linear regression method and a deep neural network (DNN). Results indicated the DNN model's superiority in accurately forecasting intradialytic hypotension during hemodialysis sessions, suggesting its potential as a management tool.

In a subsequent study, Lee et al. (2021) focused on Real-Time Prediction of Intradialytic Hypotension using a Deep Learning Model. Their research evaluated various models, including Random Forest, gradient boosting, and logistic regression. Notably, the Random Forest model demonstrated outstanding performance, achieving the highest Area Under the Curve (AUC) value of 0.812 among the tested models, exhibiting promising capabilities in intradialytic hypotension prediction.

Building upon prior research, Kim et al. (2022) undertook Predicting Intradialytic Hypotension Without Privacy Infringement, employing convolutional neural networks (CNNs) alongside regression, random forest, and XGBoost models. Notably, the CNN model exhibited promising capabilities in real-time monitoring and management of intradialytic hypotension, potentially mitigating associated risks.

In a parallel investigation, Lee et al. (2023) delved into the Prediction of Intradialytic Hypotension using pre-dialysis features, employing a deep learning-based artificial intelligence (AI) model alongside regression, random forest, XGBoost, and a one-dimensional convolutional neural network (1D-CNN). Their findings underscored the efficacy of the deep learning-based AI model in accurately predicting IDH by leveraging pre-dialysis features, thereby facilitating pre-emptive adjustments to hemodialysis settings for vulnerable patients.

Furthermore, Yun et al. (2023) explored the application of an explainable deep learning model for Real-time dual prediction of intradialytic hypotension and hypertension, utilising models such as Temporal Fusion Transformer (TFT), Recurrent Neural Network (RNN), and Light Gradient Boosting Machine (LightGBM). Their study showcased the effectiveness of the TFT-based model in simultaneous prediction of intradialytic hypotension and hypertension, thereby advancing real-time predictive capabilities.

Lastly, Vaid et al. (2023) investigated the utility of Deep Learning on Electrocardiograms for Prediction of In-hospital Intradialytic Hypotension in Patients with End-Stage Kidney Disease (ESKD). Employing a Densenet-201 model pretrained on publicly available datasets, their study demonstrated moderate accuracy in predicting which patients would develop IDH at their next hemodialysis treatment, thereby contributing to risk assessment and management strategies in clinical settings.

E. Discussion

The advent of ML and DL in healthcare has ushered in a new era of precision medicine, particularly in managing IDH. A Survey literature review was conducted to discern the efficacy of various ML and DL models in predicting IDH among hemodialysis patients. A meticulous screening was undertaken from an initial pool of 270 articles retrieved from PubMed, Google Scholar, and other relevant databases. After title screening, 200 articles were excluded, followed by the exclusion of 43 articles post-abstract screening. This left 27 articles for full-text examination. Subsequently, 12 studies were deemed pertinent and included in the final analysis. This review aims to identify which ML model most effectively predicts IDH, thereby enhancing patient outcomes in hemodialysis.

Most reviewed articles employed ML models, with fewer studies using DL models. Among the ML models, common techniques included Logistic Regression, Decision Trees, Random Forests, Support Vector Machines, and Artificial Neural Networks (Gómez-Pulido et al., 2021; Zhang et al., 2023; Liu et al., 2021). Logistic regression, Decision Trees, Random Forest, and Gradient Boosting were widely used in various studies. Logistic Regression, for example, was utilised in the study by Hong et al. (2023) to predict the occurrence of IDH, with a reported accuracy of 83.7%. Decision Trees, used in the study by Gómez-Pulido et al. (2021), achieved an accuracy of 90% in predicting the risk of IDH. Random Forest, as reported in the study by Hong et al. (2023), was the best-performing ML model, with an accuracy of 92.5%. This model effectively predicted IDH risk, as indicated by a high area under the curve (AUC) of 0.904.

VI. Challenges in the Field of Machine Learning for Predicting Intradialytic Hypotension Are

1) Data quality and availability: Many studies emphasise the need for large, high-quality datasets containing detailed patient information and real-time intradialytic measurements to train effective machine learning models. Acquiring such datasets is challenging due to privacy issues and the necessity for data integration across different healthcare systems [14].

2) Defining Intradialytic Hypotension (IDH): There is no consensus on the exact definition of IDH, with studies using varying thresholds for systolic blood pressure (SBP) decrease, nadir SBP, and mean arterial pressure (MAP) decrease. This inconsistency complicates comparing and validating different predictive models [4].

3) Model interpretability: Despite the promising performance of deep learning models like convolutional neural networks and recurrent neural networks in predicting IDH, they are often criticised for their "black box" nature, which limits interpretability. Enhancing the interpretability of these models is essential for clinical acceptance and trust [4].

4) Real-time prediction: A major challenge is to predict IDH events sufficiently in advance (e.g., 15-75 minutes) to allow for timely interventions. Most current models focus on predicting the risk of IDH throughout the dialysis session rather than providing real-time predictions [25].

5) Generalizability: Many studies develop and validate their models using data from a single centre or specific population. The generalizability of these models to diverse patient populations and dialysis settings remains uncertain.

While the Open Research Questions are:

1) Can we establish a standard definition of IDH that incorporates physiological parameters and clinical outcomes?

2) How can we enhance the interpretability of deep learning models for IDH prediction while maintaining their accuracy?

3) What is the optimal lead time for predicting IDH events to allow effective interventions, and how can we develop models to achieve this?

4) Can we create federated learning or privacy-preserving techniques to enable data sharing and integration from multiple healthcare systems while protecting patient privacy?

5) How can we validate and improve the generalizability of IDH prediction models across various patient populations, dialysis settings, and geographic regions?

6) What are the most effective interventions for preventing or mitigating IDH events once they are predicted, and how can we incorporate them into clinical workflows?

Addressing these challenges and research questions is important for developing robust, reliable, and clinically actionable machine learning models to predict intradialytic hypotension and improve patient outcomes [4] [13].

VII. Conclusion and Recommendations

Machine learning models for predicting intradialytic hypotension (IDH) in patients undergoing maintenance hemodialysis show promising results. Several articles have investigated the application of machine learning algorithms, including Decision Trees, Support Vector Machines (SVM), Random Forest, LightGBM, and custom models, for predicting IDH.

The articles demonstrated that machine learning models can achieve good predictive performance regarding accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC). Decision Trees and SVM models were commonly used and showed favourable results. Notably, models developed by Zhang et al. and Liu et al. exhibited excellent predictive performance, surpassing other models in accuracy and AUC-ROC.

Feature selection techniques, such as correlation analysis, recursive feature elimination, and principal component analysis, were employed to identify relevant predictors for IDH. Optimal

feature selection and model parameter optimisation enhanced the predictive accuracy of machine learning models.

It is important to note that the performance of machine learning models may vary depending on the dataset characteristics, feature selection methods, and model optimisation approaches employed in different studies.

Applying machine learning models for predicting IDH opens avenues for pre-emptive interventions to prevent this complication in dialysis patients. Early identification of patients at risk of IDH can facilitate targeted interventions and improve patient outcomes.

Future research should focus on large-scale studies with standardised data collection and validation of machine learning models in diverse patient populations to further advance the field. Additionally, integrating real-time physiological data from monitoring devices into machine learning models may enhance their accuracy and clinical utility.

In conclusion, using machine learning models holds promise for predicting IDH in patients undergoing maintenance hemodialysis. These models can potentially improve risk stratification and guide proactive interventions to mitigate the occurrence of IDH during dialysis treatments.

REFERENCES

[1] Chou, J.A., K. Kalantar-Zadeh, and A.T. Mathew. A brief review of intradialytic hypotension with a focus on survival. in Seminars in dialysis. 2017. Wiley Online Library.

[2] Cedeno, S., et al., Intradialytic hypotension definitions with mortality prediction capacity in a cohort of haemodialysis patients. Nefrología (English Edition), 2020. 40(4): p. 402-412.

[3] Hamrahian, S.M., et al., Prevention of Intradialytic Hypotension in Hemodialysis Patients: Current Challenges and Future Prospects. International Journal of Nephrology and Renovascular Disease, 2023: p. 173-181.

[4] Kanbay, M., et al., An update review of intradialytic hypotension: concept, risk factors, clinical implications and management. Clinical Kidney Journal, 2020. 13(6): p. 981-993.

[5] What Is Artificial Intelligence? Definition, Uses, and Types. 2020 [cited 2024 2024-5-11]; Available from: https://www.coursera.org/articles/what-is-artificial-intelligence.

[6] Kanade, V. What Is Machine Learning? Definition, Types, Applications, and Trends. 2022 [cited 2024 2024-5-11]; Available from: https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-ml/.

[7] Deng, L. and D. Yu, Deep learning: methods and applications. Foundations and trends® in signal processing, 2014. 7(3–4): p. 197-387.

[8] Schmidhuber, J., Deep learning in neural networks: An overview. Neural networks, 2015. 61: p. 85-117.

[9] Lim, W.Y.B., et al., Federated learning in mobile edge networks: A comprehensive survey. IEEE Communications Surveys & Tutorials, 2020. 22(3): p. 2031-2063.

[10] LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. nature, 2015. 521(7553): p. 436-444.

[11] Burlacu, A., et al., Challenging the supremacy of evidence-based medicine through artificial intelligence: the time has come for a change of paradigms. Nephrol Dial Transplant, 2020. 35(2): p. 191-194.

[12] Esteva, A., et al., A guide to deep learning in healthcare. Nat Med, 2019. 25(1): p. 24-29.

[13] Kim, H.W., et al., Deep learning model for predicting intradialytic hypotension without privacy infringement: a retrospective two-center study. Frontiers in Medicine, 2022. 9: p. 878858.

[14] Lee, H., et al., Deep learning model for real-time prediction of intradialytic hypotension. Clinical Journal of the American Society of Nephrology, 2021. 16(3): p. 396-406.

[15] Zhang, H., et al., Real-time prediction of intradialytic hypotension using machine learning and cloud computing infrastructure. Nephrology Dialysis Transplantation, 2023. 38(7): p. 1761-1769.

[16] Nahm, F.S., Receiver operating characteristic curve: overview and practical use for clinicians. Korean journal of anesthesiology, 2022. 75(1): p. 25.

[17] Bammani, M.S., Efficiency of LightGBM technique in Bankrupty Prediction using Polish dataset. 2020, SRH HOCHSCHULE HEIDELBERG.

[18] Vito, D. Machine Learning Techniques to Predict Intradialytic Hypotension: Different Algorithms Comparison on Unbalanced Data Sets. in Proceedings of International Conference on Data Science and Applications: ICDSA 2022, Volume 1. 2023. Springer.

[19] Dong, J., et al., Machine learning-based intradialytic hypotension prediction of patients undergoing hemodialysis: A multicenter retrospective study. Computer Methods and Programs in Biomedicine, 2023. 240: p. 107698.

[20] Gómez-Pulido, J.A., et al., Predicting the appearance of hypotension during hemodialysis sessions using machine learning classifiers. International Journal of Environmental Research and Public Health, 2021. 18(5): p. 2364.

[21] Daqing Hong 1, H.C., Xin He 1 3, Ya Zhan 1 3, Rongsheng Tong 2, Xingwei Wu 2, Guisen Li 1, Construction of an Early Alert System for Intradialytic Hypotension before Initiating Hemodialysis Based on Machine Learning. Pubmed, 2023.

[22] Chen, J.-B., et al., Deep learning for intradialytic hypotension prediction in hemodialysis patients. IEEE Access, 2020. 8: p. 82382-82390.

[23] Lee, H., et al., Prediction of intradialytic hypotension using pre-dialysis features—a deep learning–based artificial intelligence model. Nephrology Dialysis Transplantation, 2023. 38(10): p. 2310-2320.

[24] Vaid, A., et al., Deep learning on electrocardiograms for prediction of in-hospital intradialytic hypotension in ESKD patients. Kidney360, 2023: p. 10.34067.

[25] Yun, D., et al., Real-time dual prediction of intradialytic hypotension and hypertension using an explainable deep learning model. Scientific Reports, 2023. 13(1): p. 18054.

An Interactive Attention-Based Approach to Document-Level Relationship Extraction

Zhang Mei, Zhao Zhongyuan, Xu Zhitong
School of Information Science and Technology, North China University of Technology, Beijing 100144, China

***Abstract*—Document-level relation extraction entails sifting through extensive document data to pinpoint relationships and pertinent event details among various entities. This process aids intelligence analysts in swiftly grasping the essence of the content while revealing potential connections and emerging trends, thus proving invaluable for research purposes. This paper puts forward a method for document-level relation extraction that leverages an interaction attention mechanism. Initially, building on an evidence-based approach for extracting relations at the document level, the interaction attention mechanism is introduced, extracting the final layer of hidden states containing rich semantic information from the document encoder. Subsequently, these concealed states are fed into a self-attention layer informed by dependency parsing. The outputs from both elements serve as distinct supervisory signals for the interactive input. By pooling these output results, it can derive context embeddings that possess enhanced representational power. Preliminarily, relation triples are extracted using the relation classifier. In conclusion, building on the preliminary relationship results, the process of relationship inference is carried out independently using pseudo-documents created from the source material and pertinent evidence. Only those relationships with a cumulative inference score that surpasses a certain threshold are regarded as the final outcomes. Experimental findings from the publicly accessible datasets indicate commendable performance.**

***Keywords*—Document-level relation extraction; interaction attention-based; the baseline model**

I. Introduction

Conventional relation extraction models typically focus on individual sentences, overlooking the subtle contextual and semantic connections that exist between sentences throughout the entire document [1]. Document-level relation extraction is about finding and understanding the connections between different parts of a document. For instance, in document-level relation extraction, a relationship between entities might span multiple sentences, making it challenging for the model to accurately extract those relationships. Fig. 1 from the DocRED dataset shows part of a document. While the relationship between "The Legend of Zelda" and "Capcom and Flagship" may seem clear, the text has many connections that require analysis. The Legend of Zelda was created by Capcom and Flagship with guidance from Nintendo. But a more thorough examination shows that Nintendo plays a big role. They guide the design choices and overall development of the game. This scenario necessitates that the model possess robust competencies in discerning underlying connections and executing reasoning in an efficacious manner [2].

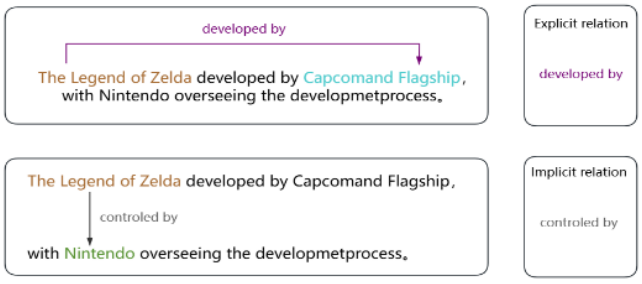


Fig. 1. Sample extraction of displayed and implicit relations under the DocRE task.

The proposed model must possess the robust capability to comprehend contextual information meticulously, refraining from concentrating on isolated sentences; it should analytically encompass the comprehensive document. The sophisticated model articulated uniquely embodies essential evidence data within the enriched supplemental information stream, utilizing this crucial input to augment relational reasoning seamlessly, thus significantly strengthening the holistic understanding of extensive documents by the model. Henceforth, numerous inadequacies remain when tackling profoundly intricate implicit connections comprehensively. Therefore, it is proposed to use interaction attention to enable the model to fully grasp contextual information, thereby further enhancing the model's reading comprehension ability for the entire document [3].

An entity in a document may be mentioned many times, but not always with the full name. It may be in the form of an abbreviation, acronym, or code name. Abbreviations and code names are analyzed on the DocRED, Re-DocRED, and CDR datasets. For example, 61.1% of the relationship instances in the same document in the DocRED dataset need to be recognized for reasoning, and only 38.9% of the relationship instances can be extracted by simple pattern recognition. This shows that commonly used pre-trained language models (e.g., Transformer, Bert, etc.) cannot completely solve the long-distance dependency and improve the overall understanding of the implicit structure of documents.

Therefore, using interaction attention to find more hidden information in the document can help document-level relation extraction models. Based on evidence-driven document-level relation extraction methods, and to mine deeper information in the document to enhance relation reasoning [4], this paper proposes using interaction attention mechanisms during document modeling to help the model uncover hidden information in the document.

II. Model Construction

To address the issues, a new document-level relation extraction model is proposed. This approach uses BERT, a pre-trained language model, with interaction attention to improve how it extracts information from data after completing evidence sentences. It also lets each attention head learn features from different parts of the document, which extends the model's capabilities. This approach helps the model learn about the whole context and makes up for the fact that attention mechanisms can only learn about what is right in front of them.

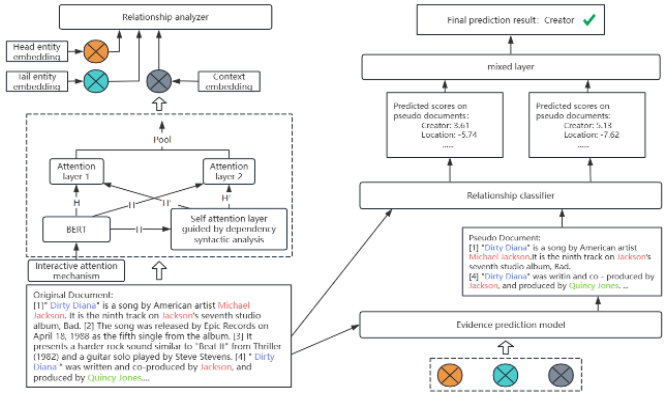


Fig. 2. IADocRE model diagram.

Fig. 2 shows how the model uses interaction attention to find information in documents. The model has two parts: relation extraction (on the left) and relation extraction and reasoning (on the right).

A. Document Level Encoder

A document, designated as d, is constituted of N sentences $\{S_n\}$, L Token tokens $\{h_l\}$, E named entities $\{e_i\}$ and all proper noun mentions $\{m_{ij}\}$ for each entity. The objective of document-level relationship extraction is to ascertain the set of potential relationships between all pairs of entities from a predefined set of relationships R, which is provided as $R \cup \{NR\}$, where NR denotes the absence of a relationship[5]. It should be noted that an entity may be referenced on more than one occasion within a single document. Consequently, for each entity e_i, there can be multiple mentions $\{ m^i_j \}$ $Ne_{ij}=1$. In the absence of a relationship between the entities in the pair (e_h, e_t), the designation NR is applied.

And make the head entity and the tail entity eh and et respectively during the test period, all entity pairs (e_h, e_t) $_{h,\ t\in(1...n),\ h\neq t}$. Essentially this is a multi-labeling classification problem because there may be multiple relationships between e_h and e_t. If the relation r exists between (e_h, e_t), it is initially classified as a positive class PT, otherwise it belongs to the negative class NT. For each entity pair of the NR relation (e_h, e_t), its evidence V_{ht} is defined as the subset of sentences in the document from which the relation can be inferred.

The document is initially encoded using BERT, which captures contextual information and semantic representations. Compared to the sentence-level encoder value focusing on the encoding of individual sentences, the document-level encoder has an inherent advantage in dealing with the task of document-level relation extraction [6]. Specifically, for a given document $d=[h_l]$, special tags are added before and after the mention of each entity using the mainstream method, i.e., [CLS] + 'Entity' + [SEP], and then encoded using an encoder:

$$H, A = \mathrm{PrLM}[h_1, ..., h_L] \quad (1)$$

where, the hidden layer state of the last layer is usually denoted as H^L, and L denotes the number of Lth layer in the BERT model.

B. Dependent Syntax Guided Self-Attention Layer

Xu et al. found that when a language model like BERT learns text, it ignores words not in the subject, predicate, or object. This research suggests that it is crucial to try to improve pre-trained language models by giving some of the attention to words other than subject, predicate, and object. The technique of dependency parsing is used to guide the model's attention towards other information in the sentence. The goal of dependency parsing is to help BERT focus on information other than the subject and predicate. In addition to using the self-attention layer to enhance attention, the incorporation of an interactive attention mechanism ensures that the enhanced model is capable of recognizing more contextual information than the original BERT model.

The text is first contextualized using the language model BERT [7]. Google found that the 12-layer Transformer encoder architecture is the best for classification. The BERT encoder captures contextual information in the text through a multi-head self-attention mechanism. The Feedforward Neural Network (FNN) performs non-linear transformations and is combined with the BERT encoder through Residual Connection and Normalization. The BERT encoder captures context in text through a self-attention mechanism. The FNN performs nonlinear transformations, which are combined and merged by residual connection and normalization. The BERT encoder was designed to improve classification effectiveness, storing rich contextual information in the document. It makes no difference if the last two or three layers are used for comparison. The last hidden layer state H is used as the initial representation of the sentence. The initial representation H of the sentence is passed to the self-attention layer, guided by the dependent syntactic analysis. The output is denoted as H'. H' shows which weights in the initial representation H are kept by the self-attentive layer. This is done according to the structure of the dependent syntactic analysis tree. Only the weights that contain dependency relations are kept. The weights of inflectional words, dummy words, and quantifiers that are not related are reduced. These can mislead the model or distract the attention weights.

Dependency parsing is about finding the links between words in a document. In Section II (B), dependency is defined as a relationship between words. Every sentence has a main word that is connected to other words.

Specifically, set the length of a sentence Sen in the document to be n, and the set of words in the sentence to be $\{W\} = \{w_1, w_2, w_n\}$, determine the ancestor set {P} of all nodes through a dependency analysis tree and create an n × n dimensional MASK matrix, denoted as M[8]. The set P consists of the directly or indirectly dominating subsets of any word w_i.

Specifically, for any word w_i ($i \in$ Sen) within a sentence, if a word w_j ($j \in$ Sen) appears in the ancestor set {P}, then the value of row i and column j of the MASK matrix M is set to 1, and all the other column positions within row i are set to 0. The computation formula is as follows:

$$M[i, j] = \begin{cases} 1 & if \quad j \in P_i \quad or \quad j = i \\ 0 & otherwise \end{cases} \tag{2}$$

In this study, the attention mechanism has three parts: the query vector Query (Q), the key vector Key (K), and the value vector Value (V), where K is usually set equal to V. The query vector and the key vector determine the weighting coefficients of the value vector. The query vector Q and the key vector K determine the weighting coefficients of V. Self-attention has the same query vector Q as the key vector K. Multi-head attention splits the query vector Q into parts and extracts multiple K-V pairs from the text. In conclusion, the above features are combined.

The final hidden layer state H of BERT is input into the multi-head self-attention layer, and the MASK matrix M of the previous complementary information is used to dot-multiply the query vector Q and the key vector K. By calculating the weights in this way, the final attention representation W_i, with i denoting each attention head, can be obtained. The computational formula is as follows:

$$A_i^{'} = soft\max(\frac{M(Q_i^{'}K_i^{'})}{\sqrt{d}}) \tag{3}$$

$$W_i^{'} = A_i^{'}V_i^{'} \tag{4}$$

The output W_i from each attention head is merged and the result is output to the FNN network, where it is activated with the GeLU. It is then passed again to another FNN network. After ths series of operations, the representation H' can be obtained following guidance by the dependency parsing after passing through the Normalization process.

C. Interactive Attention Layer

The initial representation H and the representation H' are obtained after being guided by dependent syntactic analysis. The two outputs are then fused to enhance understanding between the two representations by the model. As shown in the figure, the initial representation H of the sentence is used as the key vector Key and value vector Value of the attention layer 1, and the representation H' guided by dependent syntactic analysis is used as the query vector Query of the attention layer 1, both of them are subjected to softmax operation, and the output result is the representation vector W_1 of the attention layer 1; similarly, the output result of the attention layer 2 is the representation W_2.The computation of the two attention layers is carried out at the same time, and the computation formula is as follows:

$$W_1 = \text{softmax}(\frac{Q_1K_1^{\mathrm{T}}}{\sqrt{d_{k_1}}})V_1 \tag{5}$$

$$W_2 = \text{softmax}(\frac{Q_2K_2^{\mathrm{T}}}{\sqrt{d_{k_2}}})V_2 \tag{6}$$

Fusing W_1 with W_2 to obtain the output result of the interactive attention layer representation W_3, representation W_3 enables the overall model to pay more attention to the semantic information in the sequence related to the current position in the output. Equation is shown in 4- 7, α is used to balance the parameters of W_1 and W_2, where the value of α is 0.5.

$$W_3 = \alpha W_1 + (1-\alpha)W_2 \tag{7}$$

The BERT pre-trained language model prescribes, before input, the need to mark specific entities or sentences with special symbols in front of them and the endings with special symbols[9], i.e., [CLS] and [SEP]. The essence of [CLS] denotes the synthesized information of the whole sentence, so only the initially vector h0 of the W_3 vector is taken out and the attention A' of the whole sentence is computed using the soft-max function:

$$A^{'} = \text{softmax}(W_3[0]) \tag{8}$$

Before the document is entered into BERT, each entity e_i is required to use the embedding of the special symbol as its mention embedding m_j^i. Subsequently, the embedding of entity e_i over all its mention embeddings is obtained by employing Log-Sum-Exp pooling:

$$e_i = \log \sum_j \exp(m_j^i) \tag{9}$$

To predict the relationships between different entity pairs, the model may need to focus on different parts of the context. To capture the contextual dependencies associated with each entity pair (e_h, e_t), its contextual embedding is computed based on the interaction attention A':

$$C_{h,t} = H^T \frac{A_h^{'} \circ A_t^{'}}{(A_h^{'})^{\mathrm{T}} A_t^{'}} \tag{10}$$

where º denotes the Hadamard product, and A_h is the head entity's attention to all the tokens in the document, obtained by leveling out the mentions of the head entity. A_t is the same. Tokens that are highly attentive to both e_h and e_t must necessarily be important to both head and tail entities, and so should have more interactions on context embedding.

D. Classification of Relationships

To predict the relationship between entity pairs (e_h, e_t), the model initially computes their context-aware representations (z_h, z_t) by combining their entity embeddings (e_h, e_t) with their context embeddings $c_{h,t}$, and then utilizes a bilinear function to compute the logit of the likelihood of the existence of a relationship $r \in R$ between e_h and e_t.

$$Z_h = \tanh(W_h e_h + W_{c_h} C_{h,t}) \tag{11}$$

$$Z_t = \tanh(W_t e_t + W_{c_t} C_{h,t}) \tag{12}$$

$$y_r = Z_h W_r Z_t + b_r \tag{13}$$

where W_h, W_t, W_{ch}, W_{ct}, W_r, b_r are learnable parameters. Since the model may have different confidence levels for

different entity pairs, an adaptive threshold loss is used, which learns a virtual relationship class TH that serves as a dynamic threshold for each entity pair:

$$y_{TH} = z_h W_{TH} z_t + b_r \tag{14}$$

In the inference process, for each tuple (eh, et, r), the predicted score is obtained:

$$S_{h,r,t}^{(O)} = y_r - y_{TH} \tag{15}$$

In conclusion, the training objective for relation extraction is defined as follows:

$$L_{RE} = -\sum_{r \in P_{h,t}^T} \log\left(\frac{\exp(y_r)}{\sum_{r' \in P_{h,t}^T \cup \{TH\}} \exp(y_{r'})}\right) - \log\left(\frac{\exp(y_{TH})}{\sum_{r' \in N_{h,t}^T \cup \{TH\}} \exp(y_{r'})}\right) \tag{16}$$

Where $P_{h,t}^T$ indicates that a relationship exists between two entities and $N_{h,t}^T$ indicates that no relationship exists between entities.

E. Relational Reasoning

If the evidence sentences contain all the relevant information, there's no need to use the entire document for relationship extraction. No system can extract 100% of the evidence without omitting some sentences. Relying on extracted evidence alone may miss important information in the document, which could affect performance. The original document and the extracted evidence are combined to get the prediction results. If there are no evidence annotations, the results can be learned by the evidence prediction model from Section III or extracted by the auxiliary experiments in Section III.

Specifically, as shown in Fig. 1, a set of relationship prediction scores are initially obtained from the original document $S_{h,r,t}^{(O)}$. Then a pseudo-document d' is constructed for each entity pair by concatenating the extracted evidence sentences V_{ht} in the order in which they appear in the original document. The prediction score of the relational extraction model for the pseudo-document is recorded as $S_{h,r,t}^{(P)}$. In conclusion, fusion results are obtained by aggregating the two sets of predictions through a hybrid layer:

$$P_{Fuse(r|e_h,e_t)} = (S_{h,r,t}^{(O)} + S_{h,r,t}^{(P)} - \tau) \tag{17}$$

τ denotes the balancing parameter of the source document and the pseudo-document. The final loss function is as follows:

$$L_{Fuse} = -\sum_{r \in R} y_r \cdot P_{Fuse(r|e_h,e_t)} + (1 - y_r) \cdot \log(1 - P_{Fuse(r|e_h,e_t)}) \tag{18}$$

If $y_r = 1$ indicates that a relationship exists between the entities. Conversely, $y_r = 0$ indicates that no relationship exists between the entities.

III. Experimental Results and Analysis

The commonly used datasets for document-level relation extraction, such as DocRED, Re-DocRED, CDR, and GDA, are employed to investigate the performance differences between the model with the introduction of the interaction attention mechanism and the baseline model.

A. Data Sets

The experiments are mainly evaluated on DocRED, CDR, GDA and Re-DocRED datasets. Since CDR and GDA are relationally extracted datasets under the medical domain, since both datasets do not have annotation information such as evidence sentences, CDR and GDA are placed in the same representation for comparison [11]. Re-DocRED and DocRED tend to be generalized domain relationally extracted datasets, both containing evidence sentence information and remote supervision data, therefore CDR and GDA datasets are set in one group for comparison and DocRED and Re-DocRED are set in one group for comparison. The statistical information of the four datasets is shown in Table I.

TABLE I. Data Set Information

Statistical information	CDR	GDA	DocRED	Re-DocRED
Training Documentation	500	23353	3053	3053
verification document	500	5839	1000	1000
test document	500	1000	1000	1000
predefined relationship	2	2	97	97
Average number of entities	7.6	5.4	19.5	19.6
Average number of sentences	9.7	10.2	8.0	8.1

B. Evaluation Indicators and Parameterization

1) Evaluation indicators: As in the previous section, Precision (P), Recall (R) and F1 values were used as evaluation metrics for the experiment, which were calculated as follows:

$$P = \frac{TP}{TP + FP} \tag{19}$$

$$R = \frac{TP}{TP + FN} \tag{20}$$

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{21}$$

In addition to the above commonly used evaluation metrics, there also exists the I_{gn} F_1 evaluation metric in document-level relational extraction. I_{gn} F_1 was proposed by Yao et al. The model learns relational facts existing in the training set during the training phase, which is shared with the validation set and the test set. Then if the model has learned certain relationship facts in the training set, it will inevitably affect the model's judgment in the validation or testing phase [12]. This approach obviously produces immeasurable interference in model performance evaluation, so in order to have a fairer evaluation

index, after removing the shared relationship facts in the training and validation sets and the test set, the performance evaluation of the document-level relationship extraction model is re-conducted.

2) *Parameter setting:* The IADocRE model is based on Pytorch and Huggfacing's Transformer implementation, and the model is encoded using bert as a pre-trained language model [9]. The experiments used AdamW as the optimizer in the DocRED experiments. During the self-training phase of evidence sentences, the learning rate was set to 5e-5, Warmup was set to 0.06, and the Dropout rate was set to 0.1. When running the whole model, the experiments were trained and evaluated on four RTX 3090 24GB GPUs. The specific experimental parameters are shown in Table II.

C. Main Experiment and Analysis of Results

1) Experimental results and analysis of DocRED and Re-DocRED datasets: The main results of the IADocRE model for the two datasets are shown in Table III, the IADocRE model achieved 65.54 and 63.76 on F_1 and I_{gn} F_1 of the DocRED dataset, and 79.43 and 79.05 on F_1 and I_{gn} F_1 of the Re-DocRED dataset, and the scores on the two datasets have exceeded those of the existing baseline models. The experimental results prove that the document-level relationship extraction method based on interactive attention proposed in this paper is effective. From the table of experimental results, it can be observed that among the two mainstream methods for Document-level Relation Extraction, the effect of the model based on the sequence method is usually superior to that based on the graph method.

Table III and Table IV shows that the IADocRE model is ahead of the existing baseline model level. The I_{gn} F_1 evaluation metrics are missing for some of the models in the experimental results because some of the models were not experimentally validated on the latest dataset at the time of publication, and if the model is reproduced and validated on the latest dataset, the experimental process is again affected by the initial parameter settings, the type and number of hardware devices, and the choice of optimizer strategy. Therefore, partial results on the Re-DocRED dataset are denoted by '-'.

TABLE II. OPTIMAL PARAMETER SETTINGS FOR THE EXPERIMENT

point	parameters	(be) worth
Train	Warmup	0.06
	lr	5e-5
	evi_thresh	0.2
	Dropout	0.1
	max_grad_norm	1.0
Fine-tune	Warmup	0.06
	lr	1e-6
	evi_thresh	0.2
	Dropout	0.1
	max_grad_norm	2.0

TABLE III. EXPERIMENTAL RESULTS OF IADOCRE MODEL ON DOCRED AND RE-DOCRED DATASETS (%)

Model		DocRED				Re-DocRED			
		Dev		*Test*		*Dev*		*Test*	
		F1	***Ign F1***	***F1***	***Ign F1***	***F1***	***Ign F1***	***F1***	***Ign F1***
Graph-based Models	AGGCN	52.47	46.29	51.45	48.89	-	-	-	-
	LSR-BERT	59.00	52.43	59.05	56.97	-	-	-	-
	GLRE-BERT	-	55.40	-	57.40	-	-	-	-
	GCGCN-BERT	57.35	55.43	56.67	54.53	-	-	-	-
	GRACR-BERT	59.73	57.85	58.54	56.47	-	-	-	-
	HeterGSAN	60.18	58.13	59.45	57.12	-	-	-	-
Transformer-based Models	BERT	54.16	-	53.20	-	-	-	-	-
	BERT-Two-Step	54.42	-	53.92	-	-	-	-	-
	HIN-BERT	56.31	54.29	55.60	53.70	-	-	-	-
	CoreBERT	57.51	55.32	56.96	54.54	-	-	-	-
	SSAN-BERT	59.19	57.03	58.16	55.84	-	-	-	-
	RSMAN-BERT	59.25	57.22	59.29	57.02	-	-	-	-
	JEREX	62.24	60.39	62.15	60.29	74.77	73.34	74.79	73.48
	ATLOP-BERT	61.01	59.11	61.30	59.31	79.29	78.32	79.46	78.52
	EIDER (Rule)-BERT	62.34	60.36	62.21	60.23	-	-	-	-
	EIDER-BERT	62.48	60.51	62.47	60.42	-	-	-	-
	DocuNET	65.25	63.22	65.26	**63.23**	78.90	78.20	78.99	78.28
	IADocRE-BERT (ours)	**65.54**	**63.76**	**65.27**	63.16	**79.43**	**79.05**	**79.38**	**79.21**

In addition to the statistics on the ternary group prediction results, the experiments also included statistics on the evidence prediction results. According to the work of Huang [6] and Xie [10] , it is known that only E2GRE and EIDER have published methods for evidence extraction results. As shown in Table IV, IADocRE's method significantly outperforms E2GRE and EIDER on BERT, the evidence sentence extraction results are improved by 5.13 and 1.56 on the validation set, and by 4.51 and 1.59 on the test set, respectively in terms of the score metrics, but, there is still a lot of room for improvement in the evidence extraction.

TABLE IV. RESULTS OF EVIDENCE EXTRACTION EXPERIMENTS OF THE IADOCRE MODEL ON THE DOCRED DATASET (%)

Model	Evi F1	
	Dev	***Test***
E2GRE-BERT	47.14	48.35
EIDER-BERT	50.71	51.27
IADocRE-BERT (ours)	**52.27**	**52.86**

2) Experimental results and analysis of CDR and GDA datasets: The main results of the IADocRE model for the two datasets are shown in Table V. There is no concept of shared relational facts in the training and validation sets of the CDR and GDA datasets, so the evaluation metrics for both datasets are only F1.The IADocRE model achieves an F1 of 78.2 for the CDR dataset, and 87.8 for the GDA dataset. The performance on the GDA dataset has already exceeded that of existing benchmark models, but there remains a gap in performance on the CDR dataset, with the score being close to that of the SAIS model. Comparison of the two models reveals that: the SAIS model focuses more on the intermediate step of supervising and enhancing the model, which is a method that can more accurately capture the relevant context and entity-type information for the combination of the entity-type information and the evidence for achieving the effect of data enhancement. The intermediate steps of the supervised training process, such as Coreference Resolution, Entity Recognition and evidence-based retrieval, are clarified to help the model learn better.

TABLE V. EXPERIMENTAL RESULTS OF IADOCRE MODEL ON CDR AND GDA DATASETS (%)

Model	CDR	GDA
BERT	65.1	82.5
LSR-BERT	65.9	82.2
DHG-BERT	65.9	83.1
SSAN-BERT	68.7	83.7
GLRE-BERT	68.5	-
ATLOP-BERT	69.4	83.9
SIRE-BERT	70.8	84.7
DocuNET-BERT	76.3	85.3
SAIS-BERT	**79.0**	87.1
IADocRE-BERT (ours)	78.2	**87.8**

However, the SAIS model has the issues of increased complexity, higher data and annotation requirements, and greater consumption of computational resources compared to the IADocRE model. In summary, IADocRE is able to extract more ternary information under the medical domain dataset.

D. Ablation Experiments

1) Experimental results and analysis of ablation of DocRED and Re-DocRED datasets:

a) Analysis of the impact of interactive attention mechanisms on model performance:

All other things being equal, the contribution of the interactive attention mechanism to the model is explored by conducting multiple experiments on pre-trained language models with and without the introduction of interactive attention, respectively. The results of the ablation experiments on the DocRED and Re-DocRED datasets are shown in Tables IV-VI. When the overall model removes the interactive attention mechanism and only uses the original BERT, there is a significant decrease in the extraction effect, with F_1 and I_{gn} F_1 decreasing by 1.73 and 1.86 on DocRED and the results on Re-DocRED decreasing by 1.79 and 1.49, respectively. The ablation experiments performed on the interactive attention show that the overall model metrics decrease the most, and it can be inferred that the interactive attention mechanism is very effective in enhancing the model extraction performance.

To explore the effect of the dependent syntactic bootstrap attention layer on the model in more detail, the dependent syntactic analysis was replaced with an ordinary self-attention layer [13]. The model's extraction performance showed a decrease of 0.77 in the F_1 metric and 1.57 in I_{gn} F_1 on the DocRED dataset, indicating that the IADocRE model relies heavily on the dependent syntactic analysis attention layer, especially when ignoring relational facts. The performance significantly drops when replaced with a standard attention layer.The F_1 metric on the Re-DocRED dataset decreases by 1.04 and the I_{gn} F_1 decreased by 0.89. Considering Tan et al.'s revision of the Re-DocRED dataset, which removed a large amount of shared relational facts, the observed decreases in metrics are within the normal range. In summary, the experiments demonstrate the effectiveness of using the interactive attention mechanism in the document-level relationship extraction task.

b) Analysis of the impact of source documents and pseudo-documents on model performance:

After the ablation experiments with the interactive attention mechanism, ablation experiments were also conducted on the pseudo-document and source document parts. To explore whether the presence of source documents and pseudo documents in the relational reasoning part helps the model improve the effectiveness of relational reasoning [14]. As shown in Table IV and VI, when the model retains the interactive attention mechanism and removes the pseudo-document and only retains the source document part for relational reasoning, there is a significant decrease in the effect. When the model retains the interactive attention mechanism and removes the source document for relational reasoning, the model's F_1 and I_{gn} F_1 metrics in DocRED are 0.65 and 0.7, respectively. Observing the experimental data of the source document and pseudo-

documents' ablation in the table. It can be seen that, by removing the pseudo-document, the model's F_1 and I_{gn} F_1 metrics in DocRED decrease by 1.12 and 1.32, respectively. The ablation experimental results for both the source and the pseudo-document show that: source document and pseudo-document have a significant decrease in their effectiveness. The results of the ablation experiments on source documents and pseudo-documents show that the model is more inclined to reason on pseudo-documents than on source documents, but it cannot completely rely on pseudo-documents for reasoning. Alternatively, the pseudo-document occupies a higher position in the model's reasoning.

2) Results and analysis of ablation experiments on CDR and GDA datasets:

a) Analysis of the impact of interactive attention mechanisms on model performance:

The experimental setup is the same as in the previous subsection, and the ablation experiments are conducted using the pre-trained language model with and without the introduction of the interactive attention mechanism, respectively. As shown in Table VII, using only BERT as the pre-trained language model decreases the extraction effect by 1.6 on the CDR dataset and 1.8 on the GDA dataset. The phenomenon suggests that the introduction of the interactive attention mechanism is effective in directing the model to focus on implicit expressions in biomedical documents, such as the roles of chemicals and diseases or the associations of genes with diseases, which are not always directly explicitly mentioned. Therefore, the interactive attention mechanism is effective in deepening the understanding of biomedical domain knowledge and context.

In order to explore the effect of the dependent syntactic analysis-guided attention layer on the overall model, this layer was replaced with the ordinary attention layer for experimental analysis. As shown in Tables IV- VII, the extraction effect on CDR and GDA decreased by 0.8 and 0.6, respectively, and this result indicates that the dependent syntactic analysis-guided attention layer helps the model to improve its comprehension of logical and causal relationships between sentences. In summary, the interactive attention mechanism is effective for the model's deep understanding of complex terminology and descriptive processes in biomedicine.

b) Analysis of the impact of source documents and pseudo-documents on model performance:

As can be seen in Tables IV- VII, retaining the source document for ablation experiments results in a decrease in CDR and GDA of 1.2 and 1.1, respectively. Retaining the pseudo-document for ablation experiments results in a decrease in CDR and GDA of 1.4 and 1.5, respectively. This suggests that the model reasoning partially relies more on the source document. The reason for this is that relational extraction in the biomedical domain often requires inference at multiple levels, e.g., understanding how a specific chemical can cause a change in disease state through a specific biological pathway may require inference across multiple parts of the document. If only pseudo-documents are retained for relational reasoning, there is a high risk of losing important information at other levels. Therefore, in biomedical domain datasets, the extraction performance is better when the model is capable of in-depth understanding and reasoning about the complex interactions between entities [15].

TABLE VI. RESULTS OF ABLATION EXPERIMENTS OF IADOCRE MODEL ON DOCRED AND RE-DOCRED DATASETS

Model		DocRED		Re-DocRED	
		F1	***Ign F1***	***F1***	***Ign F1***
$IADocRE_{all}$		**65.54**	**63.76**	**79.43**	**79.05**
$IADocRE_{w/o-IA}$		63.81	61.90	77.64	77.56
$IADocRE_{w/o-DP}$		64.77	62.19	78.39	78.16
Document	-NoPseudo	64.42	62.44	78.87	78.41
	-NoOrigdo	64.89	63.06	79.02	78.92
	-all	64.40	62.38	78.94	78.50

TABLE VII. RESULTS OF ABLATION EXPERIMENTS OF THE IADOCRE MODEL ON CDR AND GDA DATASETS

Model		CDR	GDA
$IADocRE_{all}$		**78.2**	**87.8**
$IADocRE_{w/o-IA}$		76.6	86.0
$IADocRE_{w/o-DP}$		77.4	87.2
Document	-NoPseudo	77.0	86.7
	-NoOrigdo	76.8	86.3
	-all	76.6	86.2

IV. CONCLUSION

When the document-level relational extraction model fills in missing evidence information, the performance shows improvement compared to the baseline model, but there is still much room for enhancement. For example, the presence of implicit information in documents requires the model to have strong reasoning capabilities. To address this problem, the introduction of the interactive attention mechanism can help the model understand the semantic and contextual information of documents from a global perspective. By combining document information and evidence information, the model can supplement entity relationship information, thus improving the accuracy and completeness of relationship extraction. The introduction of the interactive attention mechanism helps the model to better understand the deep logical relationships, which improves the performance of the model in relational reasoning.

ACKNOWLEDGMENT

This work was supported by the Humanities and Social Science Foundation of the Ministry of Education (Grant No.21YJA740052) and National Natural Science Foundation of China (62476007).

REFERENCES

[1] Nanyun Peng, Hoifung Poon, Chris Quirk, Kristina Toutanova, and Wen-tau Yih. 2017. Cross-sentence n-ary relation extraction with graph LSTMs. Transactions of the Association for Computational Linguistics, 5:101-115.

[2] Patrick Verga, Emma Strubell, and Andrew McCallum.2018. Simultaneously self-attending to all mentions for full-abstract biological

relation extraction. in Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, V olume 1 (Long Papers), pages 872-884, New Orleans, Louisiana. association for Computational Linguistics.

[3] Ma Y, Wang A, Okazaki N. DREEAM: Guiding Attention with Evidence for Improving Document-Level Relation Extraction[J]. arXiv preprint arXiv:2302.08675, 2023.

[4] Kevin Huang, Peng Qi, Guangtao Wang, Tengyu Ma, and Jing Huang. 2021a. Entity and evidence guided document-level relation extraction. in Proceedings of the 6th Workshop on Representation Learning for NLP (RepL4NLP-2021), pages 307-315, Online. .

[5] Wenxuan Zhou, Kevin Huang, Tengyu Ma, and Jing Huang. 2021. Document-level relation extraction with adaptive thresholding and localized context pooling. In Proceedings of the AAAI Conference on Artificial Intelligence.

[6] Yiqing Xie, Jiaming Shen, Sha Li, Y uning Mao, and Jiawei Han. 2022. eider: empowering document-level relation extraction with efficient evidence extraction and inference-stage fusion. In Findings of the Association for Computational Linguistics: ACL 2022, pages 257-268, Dublin, Ireland. Association for Computational Linguistics.

[7] Benfeng Xu, Quan Wang, Y ajuan Lyu, Y ong Zhu, and Zhendong Mao. 2021. Entity structure within and throughout: modeling mention dependencies for document -In Proceedings of AAAI.

[8] Robin Jia, Cliff Wong, and Hoifung Poon. 2019. document-level n-ary relation extraction with multi-scale representation learning. in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics.

[9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. in Proceedings of NAACL.

[10] Shuang Zeng, Runxin Xu, Baobao Chang, and Lei Li. 2020. double graph based reasoning for document-level relation extraction. in Proceedings of EMNLP.

[11] Ding Xiao. Research on document-level relationship extraction technique[D]. Strategic Support Forces Information Engineering University,2023.DOI:10.27188/d.cnki.gzjxu.2023.000018.

[12] Zhu Taojie. Research on Document-Level Relationship Extraction Techniques for Long-Range Entities[D]. Strategic Support Forces Information Engineering University,2023.DOI:10.27188/d.cnki.gzjxu.2023.000079.

[13] Xiao C, Yao Y, Xie R, et al. Denoising relation extraction from document-level distant supervision[J]. arXiv preprint arXiv:2011.03888, 2020.

[14] Zhu Taojie, Lu Jicang, Zhou Gang, et al. A review of research on document-level relationship extraction techniques[J]. Computer Science,2023,50(05):189-200.

[15] Yang Yu. Research on document-level relationship extraction technique[D]. Harbin Institute of Technology,2021.DOI:10.27061/d.cnki.ghgdu.2021.002868.

Computer Modeling of the Stress-Strain State of Two Kvershlags with a Double Periodic System of Slits Weighty Elastic Transtropic Massif

Tursinbay Turymbetov[1], Gulmira Tugelbaeva[2], Baqlan Kojahmet[3],
Bekzat Kuatbekov[4], Serzhan Maulenov[5], Bakhytzhan Turymbetov[6], Mukhamejan Abdibek[7]
International University of Tourism and Hospitality, Turkistan, Kazakhstan[1]
Military of the Ground Forces named after S. Nurmagambetov, Almaty, Kazakhstan[2]
Institute of Mechanics and Engineering Science academician U. A. Dzholdasbekov, Almaty, Kazakhstan[2]
Nazarbayev Intellectual Schools – Chemistry and biology in Turkistan, Turkistan, Kazakhstan[2]
Nazarbayev University, Astana, Kazakhstan[2]
Turkestan Higher Multidisciplinary Craft College, Turkistan, Kazakhstan[3, 6]
School #26 named after Aiteke Bi, Turkistan, Kazakhstan[4, 5, 7]

***Abstract*—This paper presents computer modeling of the stress-strain state of two kvershlags with a double periodic system of slits weighty elastic transtropic massif. It introduced key concepts such as 'kvershlag', a term used to describe perpendicular cavities in a layered massif, and 'weighted elastic transtropic massif', which refers to a specialized geological structure considered in the study. These terms are critical for understanding the modeling approach. Due to the complexity of the analytical solution of this class of problems, a numerical method is used. Such a mixed problem is provided to obtain a solution by bringing it to an equivalent environment. To solve such a mixed problem, it is offered to get a solution by bringing it to an equivalent climate in terms of stiffness. The finite element method was used to solve the problem. A software package has been created to solve the stress-strain state of the two kvershlags. To ensure the correctness of the software complex, it was checked using test tasks. To study the stress-strain state of kvershlags in a weighted massif. The basic systems of equations are obtained. Algorithms are constructed and the program complex FEM_3D for solving finite element method problems is compiled. Mixed problems of the stress-strain state of cavities are solved approximately. The results of complex computer calculations are systematized, analyzed, specific conclusions are drawn and recommendations for their practical application are proposed. A computer simulated the stress-strain state of two kvershlags. The numerical solution to the given problem was obtained using the software. Results demonstrate that our numerical method approach results in 0.01%.**

***Keywords*—Transtropic; cavities; stress-strain state; deformation; finite element; slits**

I. Introduction

The first results in determining the deformation modulus of rocks were obtained in laboratory conditions. Extensive research in this direction has been undertaken in engineering schools in Germany, Austria, and Switzerland. It is enough to say the works of M. Bauschinger [1], K. Bach [2], O. Graf [3], O. Müller [4], O. Fleischer [5], K. Stöcke [6], and others. Similar works have been carried out since the 1950s under the leadership of B.V. Zalesski [7], B.P. Belikov [8] at the Donetsk Polytechnic Institute, Research Institute of G.N. Kuznetsov's [9] provided. The gap in previous research lies in the limited focus on the stress-strain behavior of kvershlags in complex geological formations. Our study addresses this by applying a numerical approach based on the finite element method to model the stress-strain state of kvershlags in a weighted transtropic massif, offering new insights into their deformation behavior. Kvershlag refers to cavities that are perpendicular to the layered massif.

Difficulties in determining the deformation of rocks cannot be solved by selecting the number of limit cracks in a given homogeneous medium. This is a three-dimensional case. Due to the variety of natural conditions, it is impossible to create a general algorithm, even in an individual case. Therefore, a convenient situation is presented below to solve the problem. To investigate the stress-strain behavior of kvershlags, a mechanical-mathematical anisotropic model with reduced moduli dependent on the slit periods and physical properties of the massif is proposed. Elasticity is treated equivalently to displacements and deformation coefficients in stones. Its solution is sought through actual experiments or modeling.

Based on thorough research, it was taken the form of layers resulting from deformable slits as a function of the medium's parameters. These layers are dependent on the medium's parameters, which have been obtained analytically. It can be related to the "deformation of layered rocks" concept.

The main concepts of the effect of slits in deformable layered rock are presented in [10-28]. Displacement vectors are considered in [29] from the perspective of the mechanics of a slit homogeneous medium. From a geological point of view, a slit can be deemed to be regarded as a space between the walls of folded rocks.

A slit in a rock zone can be called a set of slits in layers. A length characterizes each slit. The slit can only be accurately measured, like longitude. The presence or absence of slit complements also distinguishes it. Accordingly, slits are

divided into open and filled. Slit fillers consist of sand, clay soils, or other components.

These layers differ in their mineral content and physical properties. The second type of complement of slits is called a latitudinal slit, which is a supplemental slit of several millimeters along the latitude.

These layers differ in their mineral content and physical properties. The second type of complement of slits is called a latitudinal slit, which is a supplemental slit of several millimeters along the latitude.

In the case of weak connection of slits, the main conclusion about slits is drawn according to the law of asymmetric distribution of the connection along the length and width.

The main contributions of this paper are as follows:

- A computer-modeled transtropic massif with doubly-periodic slit systems as a stiffness-equivalent homogeneous continuous transtropic body.
- Elastic constants are given as functions of geometric and physical parameters of rocks and slits.
- To investigate the elastic state of two kvershlags in a weighted transtropic massif, the representation of the primary systems of solving equations of the finite element method.
- We obtained multivariate calculation results on the distribution patterns of initial stresses near two shallow kvershlags in a rock massif. These patterns depend on various factors, including the parameters of the slits and other initial data.

The rest of this paper is organized as follows:

Section II presents the state of the art in studying the stress-strain state of shallow underground cavities in solid and slotted weakened massifs and discusses methods for analyzing the state of weighted rock strata with two treks—Problem Statements. Section III presents our proposed finite element approach using isoparametric computational elements. A transtropic massif with double periodic slit systems is replaced by a continuous transtropic body equivalent in stiffness to the primary medium by solving the reduction problem. In Section IV we determined the elastic constants of a transtropic body that is equivalent in stiffness to the main massif with slits. The elastic constants depend on the elastic properties of the main massif and the geometry of the slits. In the values of the initial elastic components of stresses and displacements in the transtropic massif around two kvershlags depending on the initial parameters by implementing the developed algorithms, computer program complexes, conducting multivariate calculations and analyzing the numerical results. Section V presents the experimental results, compares the performance of our proposed method with existing state-of-the-art techniques, and discusses the method with existing state-of-the-art techniques and the implications of our findings. Finally, Section VI summarizes the results and suggests directions for future research.

II. Related Works

In the last century, the works of foreign scientists mainly involved theoretical research on the stress-strain state of underground cavities in the isotropic massif. Using the symmetry of the biharmonic solutions and based on the unique properties of harmonic functions, O.Müller and K. Stocke reviewed the relevant class of problems. G.V. Kolosov, N.I. Muskhelishvili [30], in the solution of plane problems of the theory of elasticity of an isotropic body, has successfully used the method of the Complex Variable Theory. W. Wittke [31] extended the anisotropic jointed rock model (AJRM) and the corresponding analysis methods to a broader spectrum of rock types. His design approach has been applied to many tunneling, dam, and slope design projects.

The analytical function proposed by Appel considers the state of one and many related isotropic bodies with a circular hole. L.A. Filshtinsky considered orthotropic structures with a doubly periodic system of circular holes [32] and a body with elliptical holes A.S. Kosmodamiansky, M.M. Neskorodev [33]. A.S. Kosmodamiansky investigated the stress-strain state of an anisotropic elastic body with three and endless rows of holes, and based on these decisions, Zh.S. Erzhanov, K.K. Kaydarov, M.T. Tusupov [34] studied the effects of the slots on the static stress state of underground workings. Zh.S. Erzhanov, Sh.M. Aytaliev, and Zh.K. Massanov [35] proposed a computational mechanics and mathematical model of the anisotropic elastic deformation of the rock mass with doubly periodic systems slots. They solved the problem by bringing the elastic constants obtained from the transtropic body, the equivalent stiffness central massif with slots, depending on the elastic properties and the geometry of the slots. Based on this model, they studied static initial elastic states, mainly single underground cavities deep foundation of rigorous and approximate methods, and subsequent creeping cavities state-based Ǝ-Algebra Operators U.N. Rabotnov [36] and the Theory of Creep of Rocks Zh.S. Erzhanov [37].

Different types of immersed tunnels (IMT) as well as their construction methods are discussed by [38-45].

These scientists have made significant contributions to the theory of the Finite Element Method and its application to solving complex problems of statics and dynamics of solid mechanics: L. Segerlind [46], B.Z. Amusin, A.B. Fadeev [47], Zh.S. Erzhanov, T.D. Karimbaev [48], A.D. Omarov, Zh.K. Massanov, N.M.Mahmetova [49], L.B. Atymtayeva, B.E. Yagaliyeva [50] and others [51-53].

A. The Task

Investigated the static stress and strain state of two kvershlags lying in a heavy transtropic massif, depending on the degree of discontinuity, conform to small sloping layers at an angle φ. Let H denote the depth of the distance between their centers two $2L$ (Fig. 1).

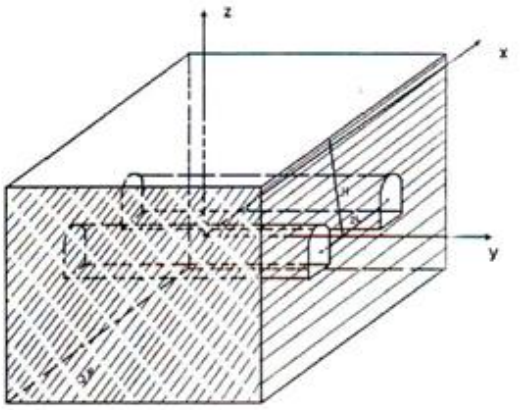


Fig. 1. Three-dimensional view.

B. The Task Explained

The plane of the cross-sectional areas with anisotropic in-plane deformation slits; efforts are at infinity (Fig. 2).

$$\sigma_x^{(\infty)} = p, \quad \sigma_z^{(\infty)} = q, \quad \tau_{xz}^{(\infty)} = r \tag{1}$$

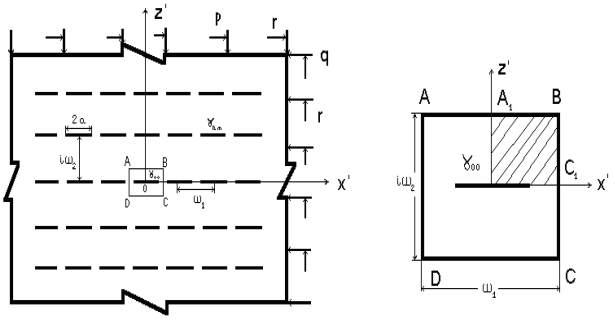


Fig. 2. Surface with a periodic system of slits.

Here γ_{00} - main crack; γ_{nm} - circuits and their length; n,m - indices, $\omega_1, i\omega_2$ - periods of slits in the directions of the axes x and z; circuits are free of external loads. E_j, ν_j, G_2 $(j=1,2)$ - elastic properties transtropic massif slots. By solving the problem of bringing to an anisotropic body with the boundary conditions (1) given elastic parameters E_i^e, ν_i^e, G_2^e, $(i=1,2)$, transtropic solid body, equivalent stiffness anisotropic massif with slits are the following formulas:

$$E_1^э = E_1, \ \nu_1^э = \nu_1, \ \nu_2^э = \nu_2,$$

$$E_2^{э^{-1}} = E_2^{-1} + 2\omega^{-1} < 2\,\mathrm{Re}\sum_{j=1}^{2} q_j \Phi_j\left(x + i\beta_j 0.5\omega, q\right) >, \tag{2}$$

$$G_2^{э^{-1}} = G_2^{-1} + 2\omega^{-1} < 2\,\mathrm{Re}\sum_{j=1}^{2}\left[p_j \Phi_j\left(x + i\beta_j 0.5\omega, r\right) + q_j \Phi_j\left(x + i\beta_j 0.5\omega, r\right)\right] >$$

Here < > - symbol averaging values, β_j – anisotropy parameters;

$$\Phi_j\left(z_j\right) = \left(B_j + iC_j\right) z_j + \sum_{k=1}^{\infty} a_{2k-1,j}\zeta_j^{-(2k-1)} + \sum_{k=1}^{\infty}\sum_{k=1}^{\infty} a_{2k-1,j} B_{jik}\left(\zeta_j^{2k-1} + \zeta_j^{-(2k-1)}\right),$$

$$B_1 = 0.5\left(p + \beta_2^2 q\right)\left(\beta_2^2 - \beta_1^2\right)^{-1} \tag{3}$$

$$B_2 = 0.5\left(p + \beta_1^2 q\right)\left(\beta_2^2 - \beta_1^2\right)^{-1}$$

$$C_1 = 0, \ C_2 = 0.5 r \beta_2^{-1};$$

III. Proposed Method

A. The Solving Problem

Hooke's law of transtropic massif with cavities with generalized plane strain relative to the Cartesian coordinate system $Oxyz$ (see Fig. 1):

$$\{\sigma\} = [\overline{D}]\{\varepsilon\}; \tag{4}$$

where $\{\sigma\} = \left(\sigma_x, \sigma_z, \tau_{yz}, \tau_{xz}, \tau_{xy}\right)^T$, $[\overline{D}] = [d_{ij}]$, $(i, j = 1,2,...,5)$; - deformation coefficients defined by the formulas [54].

Here $E_k^e, \nu_k^e, G_2^e (k=1,2)$ - effective elastic constants transtropic massif equivalent stiffness transtropic massif with slits, which depends on the elastic constants of the last and the geometry of the slits $a, \omega, i\omega$.

B. The Use of Numerical Methods

The cross-section in plane ABCD kvershlag planes of deformation using units to isoparametric calculation elements (Fig. 3).

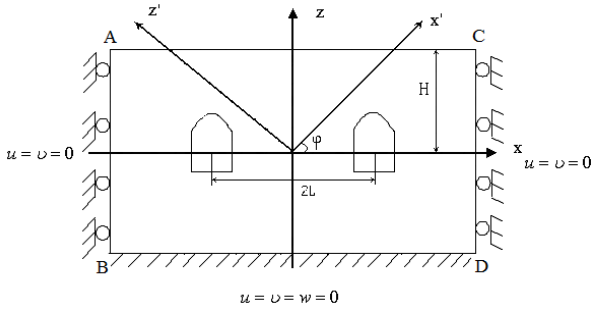


Fig. 3. Two-dimensional view.

Constitute the essential resolution of the system of algebraic equations finite element method's three $3n$-order relative to the projections of moving points, and it can be solved with the following boundary conditions [55]:

Base BD calculation area ABCD non-deformable –

$$u = v = w = 0; \tag{5}$$

Sides AB and CD under the weight of rocks moved only in the vertical direction due to a lack of influence of cavities –

$$u = v = 0, \quad w = w(z). \tag{6}$$

The study estimated the area with cavities is automatically split into isoparametric elements using the object-oriented program FEM_3D (Fig. 4).

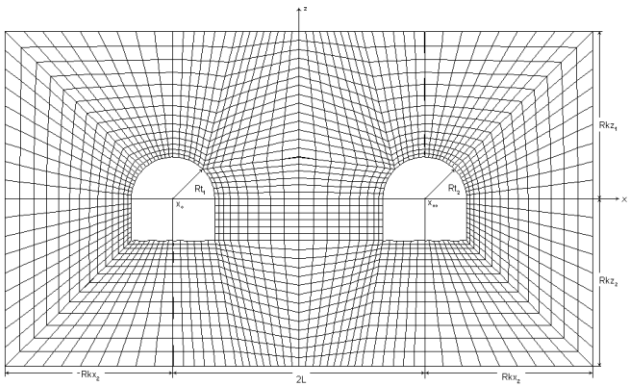
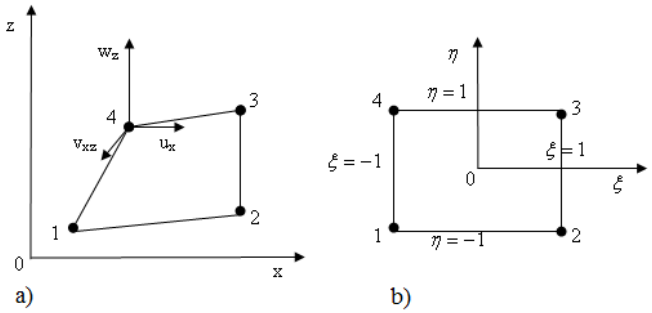


Fig. 4. A layout of the estimated area for isoparametric elements.

In solving the problem, we considered a generalized plane calculation algorithm for points isoparametric element (Fig. 5).



a) Cartesian coordinate system xoz;
b) Local coordinate system $\xi o \eta$;

Fig. 5. Four points isoparametric elements.

Each point acts as the vertical force of the weight:

$$f_{z_i} = -\frac{\gamma S}{4};\ (i = 1,2,3,4) \tag{6}$$

To solve the fundamental system of equations with the Finite Element Method's displacement components with the boundary conditions (5), (6) rigorous methods are complex; therefore, it can be solved in an iterative method of Gauss-Seidel-relaxation factor with a given accuracy [56]:

$$\{F\} = [K]\{U\} \tag{7}$$

here $[K] = \sum_{i=1}^{n} [k^e]_i$ – stiffness matrix of the system;

$\{U\} = (u_1,...,u_R, w_1,...,w_R, v_1,...,v_R)^T$ – displacement vector;

$\{F\} = (F_{x_1},...,F_{x_R}, F_{z_1},...,F_{z_R}, F_{y_1},...,F_{y_R})^T$ – force vector.

An attractive feature of this method is as follows: the stiffness matrix in this method is prepared only once and can be used for iterating both its own elements and the column elements of the matrix. During the first iteration, the value of the matrix is unknown, but in subsequent iterations, the value of the previous iteration is used.

IV. Experimental Results

Thus, the searched $E_1^э, E_2^э, G_2^э, \nu_1^э, \nu_2^э$ parameters describe the anisotropic medium weakened by two periodic slits according to the formulas (2) and (3). Elastic and geometrical parameters of slits shape them. It has been shown above that the equivalent homogeneous folding condition for anisotropy is realized by introducing the measure of elastic parameters. Below (Table I) is the elasticity parameter of anisotropic rocks weakened by two periodic slits [48].

Table I gives elastic parameters and their anisotropy parameters for anisotropic siltstone [35], and several cases are given by entering different values of relations of geometrical parameters.

TABLE I. Value of Equivalent Elastic Properties and Anisotropy Parameters

$\frac{\omega}{a}$	Given elastic parameters					Anisotropy parameters		
	$E_1^э \cdot 10^{-4}$ (MPa)	$E_2^э \cdot 10^{-4}$ (MPa)	$G_2^э \cdot 10^{-4}$ (MPa)	$\nu_1^э$	$\nu_2^э$	$k^э$	$n^э$	$l^э$
Anisotropic siltstone [48]								
∞	1,074	0,523	0,120	0,413	0,198	1,56	3,64	1,78
6,0	1,074	0,426	0,097	0,413	0,198	1,73	4,02	1,98
4,0	1,074	0,311	0,073	0,413	0,198	2,03	4,59	2,28
3,0	1,074	0,214	0,056	0,413	0,198	2,45	5,22	2,60
2,5	1,074	0,148	0,045	0,413	0,198	2,95	5,82	2,90

$\frac{\omega}{a} = \infty$ The given values correspond to the limit values of the layer. The first value represents the scenario with no slits, while the second value indicates that the connection between the layers is discontinuous.

Table II presents the verification of the correct operation of the developed algorithms and software systems to solve the test problem of the elastic stress state circular cavity in an anisotropic massif with the horizontal plane of isotropy in the plane strain and hydrostatic stress distribution in a pristine environment. Because of the symmetry of the problem, a quarter of the area of the cavity is divided into 342 isoparametric elements with the help of 380 points. The basic system of equations is solved in about 1140 with 1000 iterations. Unlike values of displacements at characteristic points of contour obtained by iterative and strict known methods, it is no more than 1-2%.

TABLE II. COMPARISON OF THE SHEAR IN THE CONTOUR POINTS OBTAINED BY ITERATION AND RIGOROUS METHODS

θ, deg	$-\sigma_\theta^{cont}/\gamma H$			
	$-\sigma_\theta^{anal}/\gamma H$ Precise method (test)	$-\sigma_\theta^{FEM}/\gamma H$ FEM	$\lvert\sigma_\theta^{anal}/\gamma H-\sigma_\theta^{FEM}/\gamma H\rvert$	$\frac{\lvert\sigma_\theta^{anal}-\sigma_\theta^{FEM}\rvert}{\lvert\sigma_\theta^{anal}\rvert}$
0	3.079	3.040	0.039	0.01
30	1.510	1.493	0.017	0.01
60	1.706	1.694	0.012	0.007
90	2.692	2.631	0.061	0.022

A. Results

In this section, we demonstrate the results that were obtained. Twelve points were obtained on the boundary of a double kvershlags in transtropic rocks weakened by weighted elastic two-period slits (Fig. 6). As a result of multivariate calculations, the change of tangential stress at the boundaries of the two kvershlags was studied. In this case, two kvershlags are located at depths of H=10M and distances between cavities of L=5M. Our findings align with recent studies (e.g., Turymbetov et al., 2020) on the stress distribution patterns in complex geological structures, but provide additional insights into the behavior of kvershlags when subjected to various stress conditions. A key difference is the extended analysis of the interaction between two closely positioned kvershlags, which has not been previously examined in detail.

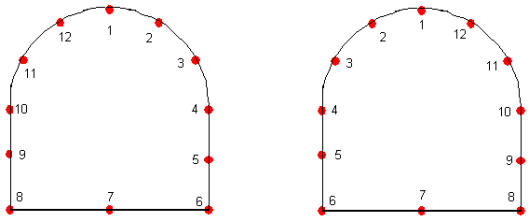


Fig. 6. Given points on the boundary of the cavity.

Tables III-V presents the variation of tangential stresses in the values w/a=∞, w/a=6, and w/a=4, the ratios of geometrical parameters.

TABLE III. THE CHANGE OF TANGENTIAL STRESS AT THE BOUNDARY OF TWO KVERSHLAGS IN THE GEOMETRIC PARAMETER OF SLITS IS W/A=∞

Left cavity					
	φ=0	φ=30⁰	φ=45⁰	φ=60⁰	φ=90⁰
1	-0,2	-0,408	-0,141	-0,166	-0,37
2	-0,6	-0,579	-0,785	-0,831	-0,79
3	-1,585	-2,366	-2,495	-2,066	-1,876
4	-2,405	-3,277	-3,396	-2,535	-2,093
5	-2,218	-3,15	-3,121	-2,335	-1,89
6	-3,33	-5,929	-4,687	-3,811	-3,514
7	0,687	0,712	0,699	0,38	0,193
8	-3,895	-5,382	-5,429	-4,668	-4,066
9	-2,082	-2,53	-2,518	-2,093	-1,767
10	-2,236	-2,83	-2,869	-2,361	-1,95
11	-1,727	-2,997	-2,845	-2,142	-1,979
12	-0,744	-2,18	-1,305	-0,712	-0,961
Right cavity					
1	-0,216	-0,54	-0,644	-0,465	-0,372
2	-0,595	-0,529	-1,301	-1,185	-0,78
3	-1,583	-2,371	-2,469	-2,003	-1,875
4	-2,376	-3,242	-3,056	-2,337	-2,085
5	-2,268	-3,263	-2,734	-2,122	-1,955
6	-3,925	-6,764	-6,717	-5,199	-4,111
7	0,555	0,557	0,29	0,192	0,17
8	-3,901	-5,291	-5,734	-4,929	-4,073
9	-2,082	-2,486	-2,447	-2,069	-1,766
10	-2,319	-2,899	-2,744	-2,157	-1,992
11	-1,727	-2,993	-2,851	-2,144	-1,976
12	-0,742	-2,18	-1,534	-0,914	-0,955

TABLE IV. THE CHANGE OF TANGENTIAL STRESS AT THE BOUNDARY OF TWO KVERSHLAGS IN THE GEOMETRIC PARAMETER OF SLITS IS W/A=6

Left cavity					
	φ=0	φ=30⁰	φ=45⁰	φ=60⁰	φ=90⁰
1	-0,197	-0,037	-0,163	-0,082	-0,359
2	-0,605	-0,224	-0,828	-0,209	-0,783
3	-1,574	-2,228	-1,654	-0,78	-1,874
4	-2,414	-3,754	-2,329	-1,417	-2,095
5	-2,235	-3,78	-2,243	-1,52	-1,892
6	-3,343	-6,045	-3,198	-2,413	-3,505
7	0,8	1,166	0,523	0,322	0,2
8	-3,924	-5,472	-4,277	-2,603	-4,05
9	-2,098	-2,808	-1,97	-1,243	-1,77
10	-2,243	-3,206	-2,182	-1,459	-1,953
11	-1,714	-3,241	-1,668	-1,329	-1,976
12	-0,747	-2,018	-0,216	-0,311	-0,953
Right cavity					
1	-0,215	-0,781	-0,57	-0,528	-0,362
2	-0,601	-0,888	-1,332	-0,836	-0,773
3	-1,573	-2,388	-1,647	-0,9	-1,873
4	-2,383	-3,459	-2,126	-1,153	-2,087
5	-2,282	-3,452	-2,006	-1,134	-1,958
6	-3,954	-8,248	-4,497	-3,614	-4,101
7	0,637	0,471	0,224	-0,066	0,177
8	-3,929	-5,963	-4,653	-3,163	-4,057
9	-2,098	-2,746	-1,962	-1,22	-1,769
10	-2,328	-3,162	-1,992	-1,144	-1,994
11	-1,713	-3,225	-1,652	-1,184	-1,973
12	-0,746	-2,183	-0,388	-0,474	-0,946

TABLE V. THE CHANGE OF TANGENTIAL STRESS AT THE BOUNDARY OF TWO KVERSHLAGS IN THE GEOMETRIC PARAMETER OF SLITS IS W/A=4

Left cavity					
	φ=0	φ=30^0	φ=45^0	φ=60^0	φ=90^0
1	-0,187	-0,011	-0,09	-0,282	-0,347
2	-0,622	-0,67	-0,218	-0,474	-0,775
3	-1,575	-2,24	-0,399	-0,403	-1,872
4	-2,406	-3,588	-1,103	-0,7	-2,097
5	-2,242	-3,42	-1,24	-0,688	-1,895
6	-3,372	-4,439	-1,231	-0,429	-3,495
7	1	1,151	0,473	0,233	0,208
8	-3,975	-5,215	-1,893	-1,683	-4,031
9	-2,105	-2,57	-0,956	-0,601	-1,773
10	-2,235	-2,951	-1,093	-0,586	-1,956
11	-1,705	-2,696	-0,772	-0,415	-1,973
12	-0,762	-1,118	0,172	0,179	-0,942
Right cavity					
1	-0,209	-0,763	-0,649	-0,599	-0,35
2	-0,619	-1,291	-1,06	-1,102	-0,765
3	-1,574	-2,169	-0,659	-0,691	-1,871
4	-2,375	-3,196	-0,826	-0,521	-2,089
5	-2,286	-3,017	-0,767	-0,267	-1,961
6	-4,013	-6,67	-2,537	-1,61	-4,089
7	0,782	0,53	-0,123	-0,116	0,185
8	-3,98	-5,582	-2,675	-2,275	-4,038
9	-2,105	-2,486	-0,954	-0,642	-1,772
10	-2,317	-2,799	-0,896	-0,617	-1,997
11	-1,705	-2,653	-0,646	-0,474	-1,971
12	-0,76	-1,329	0,05	0,011	-0,936

V. DISCUSSION

Despite certain achievements in studying the state of individual deep cavities using isotropic and anisotropic elastic computational models of rock strata, there has been no systematic study of the stress-strain state of two drifts with arbitrary cross-sectional shapes and depths of location in a large inclined layered massif with a system of slits under conditions of elastic deformation of rocks.

The study of the regularity of distribution of elastic stresses and displacements in the vicinity of cavities of arbitrary depth and cross-sectional shapes depending on the inhomogeneous fractured structure is not only of theoretical interest but also has direct practical significance.

This study systematically investigates the distribution patterns of elastic stresses and displacements near two kvershlags of arbitrary profile shape and depth using the finite element method. The study is based on a homogeneous transtropic computer-modeled inclined-fine-layered massif with a double periodic system of slits, and is conducted under conditions of generalized plane deformation.

VI. CONCLUSION

Future research could focus on refining the numerical model by incorporating real-time field data, which could improve the accuracy of predictions regarding the stress-strain behavior of kvershlags in various geological settings. Additionally, the extension of this model to analyze other types of underground structures could further broaden its application.

In this paper, we presented a finite-element computational model of anisotropic rocks with a double periodic system of slits and elastic properties in the form of a transtropic body with given moduli. To ensure the generalizability and adaptability of the model, we also performed multivariate calculations.

Using an accepted model of deformation for the weighted massif, we formulated a boundary problem to determine the initial elastic stress-strain state of the kvershlag cavity. We obtained the basic equation system for solving this problem using the finite element method.

A calculation algorithm and a program complex for studying the elastic state of double kvershlags of arbitrary depth have been developed.

We conducted numerical calculations and analyzed the influence of geometric and physical parameters, as well as rock elasticity properties, on the components of stresses and displacements in the vicinity of cavities. Based on these analyses, we arrived at specific conclusions.

Our analysis revealed that as the slits approach each other, the value of displacements increases, and the value of stresses changes. Additionally, we observed that the numerical value of stresses and displacements increases as the cavity location deepens and the distance between them decreases.

REFERENCES

[1] Bauschinger M. (1884). Versuche über Elastizität und Festigkeit. München (Vol. 10, pp. 50-68).

[2] Bach K., Baumann R. (1924). Elastizität und Festikgiet. Berlin, 687 p. Springer.

[3] Graf O. (1926). Versuche über die Druckelastizität von Basalt, Gneis, Muschelkaik, Quarzit, Granit, Bundsandstein, sowie von Hochofenstuckshlake. Beton und Eisen. (Vol. 22, pp. 399-406).

[4] Müller O. (1930). Untersuchungen an Karbongesteinen zur Klärung von Gebirgsdruckfragen. Glückauf (Vol. 47, pp. 1601-1612).

[5] Fleischer O. (1934). Bergmännische Beobachtungen und Versuche über Gebingsbewegungen im obershlesischen beim. Glückauf (Vol. 31, pp. 716-721).

[6] Stöcke K. (1937). Für das Gebingsdruckproblem wichtige Begriffe aus der technischen Mechanik. Zeitschrift für des Berg-Hütten und Salinenwesen. Bd. 84. (Vol. 11, pp. 465-467).

[7] Zalesskiy B.V., Florenskiy K.V. (1952). About some basic prerequisites for testing stone building materials for frost resistance. Proceedings of the Institute of Geological Sciences (pp. 39-50).

[8] Belikov B.P. (1952). Elasticity moduli of different types of rocks of the USSR. Proceedings of the Institute of Geological Sciences (pp. 3-38).

[9] Kuznetsov G.N. (1947). Mechanical properties of rocks. M: Ugeltekhizdat, 180 p.

[10] Krieger N.I. (1951). Cracking and methods of its study at hydrogeological survey. Materials on engineering geology, (Vol. 2, pp. 7-139).

[11] Kuvaev N.N. (1960). On the persistence of crack orientation at different depths. Works of VNIMI, (Vol. XXXIV, pp. 305-307).

[12] Protodyakonov M.M., Chirkov S.E. (1964). Cracking and strength of rocks in the massif. M: Nauka, 67 p.

[13] Laufer H. (1958). Gebirgsklassifizierung für den Stollenbau. Geologie und Bauwesen (Vol. 1, pp. 46-51).

[14] Belousov V.V. (1952). Tectonic ruptures, their types and mechanism of formation. Proceedings of the Geophysical Institute of the USSR Academy of Sciences (pp. 1-145).

[15] Mikhailov A.E. (1956). Field methods of studying cracks in rocks. M: Gosgeoltekhizdat, 132 p.

[16] Permyakov E.N. (1949). Tectonic fracturing of the Russian platform. Materials to the knowledge of the geological structure of the USSR. M: 210 p.

[17] Chernyshev V.F. (1965). About regularities in the position of spalling and fracturing cracks. Proceedings of the Institute of Geological Sciences of the Academy of Sciences of the USSR, (Vol. 162, pp. 146-150).

[18] Kushnarev I.P., Lukin L.I. (1960). On the study of fracture tectonics. Problems of Tectonophysics. M: Gosgeoltechizdat, (pp. 99-110).

[19] Neishtadt L.I. (1957). Methods of geological study of rock fracturing in engineering-geological research. M: Gosenergoizdat, 103 p.

[20] Sokolov D.S. (1962). Basic conditions of karst development. M: Gosgeoltekhizdat, 322 p.

[21] Rats M.V., Chernyshev S.N. (1970). Cracking and properties of fractured rocks. M: Nedra, 159 p.

[22] Nevskiy V.A. (1964). Genetic classification of rock cracks. Investigations of universities, Geology and Exploration, (pp. 3-13).

[23] Belousov V.V. (1962). Basic questions of geotectonics. M: Gosgeoltekhizdat, 608 p.

[24] Shultz S.S. (1964). On the study of planetary fracturing. International Geological Congress, XXII session. Problems 4. Reports of Soviet geologists to the XXII session of the IGC. M: Nauka, (pp. 147-152).

[25] Savarensky F.P. (1939). Engineering Geology. M: State Scientific and Technical Publishing House, 488 p.

[26] Kolomensky N.V. (1956). Engineering Geology. M: Gosgeoltekhizdat, 320 p.

[27] Kirillova I.V. (1945). Experience in analyzing fracturing in layered strata of the eastern dip of the Trialet folded region. DAN USSR. pp. 389-892.

[28] Skaryatin V.D. (1965). Classification of fractures and application of aerial photography and some other methods in the study of rock fracturing. Proceedings of the 2nd All-Union Meeting on fractured oil and gas reservoirs. M: Nedra, (pp. 66-80).

[29] Barenblatt G.I. (1961). Mathematical theory of equilibrium cracks formed at brittle fracture. Journal of Applied Mechanics and Technical Physics. M: (pp. 3-56).

[30] Muskhelishvili N.I. (1966). Some basic problems of the mathematical theory of elasticity. M: Science, 707 p.

[31] Wittke W. (2014). Rock Mechanics Based on an Anisotropic Jointed Rock Model (AJRM). Germany, 891 p.

[32] Filshtinsky L.A. (1967). Stress in the correct doubly periodic grids (lattices). Engineering journal. Mechanics of Solids (Vol. 1, pp. 20-28).

[33] Kosmodamianski A.S., Neskorodev M.M. (1970). Doubly periodic problem for an anisotropic structure weakened elliptical holes. "Reports of the Academy of Sciences URSR Ser. A (Vol. 7, pp.31-42).

[34] Erzhanov Zh.S, Kaidarov K.K., Tusupov M.T. (1969). Mountain range with a discontinuous layer coupling (plane problem)."Mechanical processes in the rock mass. Alma-Ata, Nauka, 189 p.

[35] Erzhanov Zh.S, Aytaliev Sh.M., Massanov Zh.K. (1971). Stability of horizontal caves in the pan-layered massif. Alma-Ata, Science, 160 p.

[36] Rabotnov U.N. (1948). Some problems in the theory of creep. Vestnik MGU, (Vol. 10, pp. 145-152).

[37] Erzhanov Zh.S. (1964). The theory of creep of rocks and its application. Alma-Ata, Nauka, 175 p.

[38] Glerum A. (1992). Options for tunneling: a personal story. Tunn. Undergr. Space Technol. 7 (4), 313–315.

[39] Rasmussen N.S. (1997). Concrete immersed tunnels — Forty years of experience. Tunn. Undergr. Space Technol. 12 (1), 33–46.

[40] Bucher C., 2006. Applications of random field models in stochastic structural mechanics. In: Advances in Engineering Structures, Mechanics & Construction, vol. 140, pp. 471–484.

[41] Soubra A.-H., Youssef Abdel Massih, D., Kalfa, M. (2008). Bearing capacity of foundations resting on a spatially random soil. GeoCongress, pp. 66–73.

[42] Li W., Fang Y., Mo H., Gu R., Chen J., Wang Y., Feng D. (2014). Model test of immersed tube tunnel foundation treated by sand-flow method. Tunn. Undergr. Space Technol. 40, 102–108.

[43] Hicks M.A., Li Y. (2018). Influence of length effect on embankment slope reliability in 3D. Int. J. Numer. Anal. Methods Geomech. 42 (7), 891–915.

[44] Müller S., Schüler L., Zech A., Heße F., 2022. GSTools v1.3: a toolbox for geostatistical modelling in python. Geosci. Model Dev. 15 (7), 3161–3182.

[45] Cornelis M.H., Oswaldo M.N., Bas J. (2024). The influence of spatial variation on the design of foundations of immersed tunneles: Advanced probabilistic analysis. Tunnelling and underground space technology, 147, 3-35.

[46] Segerlind L. (1979). Application of the Finite Element Method. Moscow, Mir, 392p.

[47] Amusin B.Z., Fadeev A.V. (1975). The Finite Element Method for solving mining mechanics. Moscow, Nedra, 142 p.

[48] Erzhanov Zh.S., Karimbayev T.D. (1975). The Finite Element Method in problems of rock mechanics. Alma-Ata, Science, 238 p.

[49] Omarov A.D., Massanov Zh.K, Makhmetova N.M. (2002). Static and seismic stress state transport of underground structures in anisotropic nonlinear geometric massif. Almaty, Bastau, 244 p.

[50] Atymtayeva L.B., Masanov Z.K., Yagaliyeva B.E. (2013). Analysis of stress-strain state of diagonal mines in multilinked anisotropic massif under the actions of elastic waves. Lecture Notes in Engineering and Computer Science, 134–138

[51] Marchandise E., Remacle J-F. (2006). A stabilized finite element method using a discontinuous level set approach for solving two phase incompressible flows. J comput phys 219 (2), 780-800.

[52] Zhao L., Mao J., Bai X., Li T., Williams JJR. Finite element implementation of an improved conservative level set method for two-phase flow. Comput & Fluids, 100, 38-54.

[53] Kardani V., Nazem M., Carter JP., Abbo AJ. (2015). Efficiency of high-order elements in largedeformation problems of geomechanics. Int. J. Geomech. 15 (6), 401-411.

[54] Turymbetov T., Azhikhanov N., Zhunisov N., Aimeshov Zh. (2015). Stress-strain state of two diagonal cavities weighty inclining layered massif system with slots in terms of elastic-creep deformations. World conference on technology, innovation and entrepreneurship. Procedia-Social and Behavioral Sciences. Istanbul, Turkey, (pp. 2263-2271).

[55] Turymbetov T., Kuatbekov B., Maulenov S., Aimeshov Z. (2019). The Stress-Strain State of the Two Shtreks a Weighty Obliquely Layered Massif System With Slits in Terms of Elastic Deformation of Rocks. 3rd World Conference on Technology, Innovation and Entrepreneurship (WOCTINE), Procedia Computer Science, Istanbul, Turkey, (Vol.158, pp. 355-360).

[56] Azhikhanov N.T., Zhumagulov B.T., Turymbetov T.A., Bekbolatov A.B. (2020). Stressed-deformed state of two drifts in a tiltly layered cracked array in the conditions of elastic deformations of rocks. Al-Farabi Kazakh National University. Journal of Mathematics, Mechanics and Computer Science. Almaty, Kazakhstan, 1 (105), 120-128.

Smart System for Driver Behavior Prediction

Hajar LAZAR, Zahi JARIR
Computer Science Engineering Laboratory, Cadi Ayyad University, Marrakesh, Morocco

***Abstract*—Driver behavior has recently emerged as a challenging topic in Traffic risk studies. Despite the advances in this topic, the challenges still remain. In fact, the current contribution deals with predicting at Real-time driver behavior based on machine learning techniques handling data sensing collected from smartphone sensors (accelerometer, gyroscope, GPS) and from OBD II. To ensure prediction at real time, we used a real-time architecture utilizing Atlas MongoDB service to synchronize data communication. Furthermore, we opt Random Forest model that demonstrates the highest performance compared to other models. This model has the advantage of predicting and preventing by warning a driver if his or her driving style is aggressive, moderate or slow. The proposed system aims to give more information about incidents to gain a better understanding of their causes.**

***Keywords*—Driver behavior prediction; OBD II; smartphone sensors; intelligent transport system; traffic safety**

I. Introduction

Traffic accidents have been on the rise over the past decade and have turned into a serious problem that affects the safety of people and their property. The World Health Organization report states [1] that pedestrians, cyclists, and motorcyclists are responsible for more than 50% of all road traffic deaths and, every year, road traffic crashes result in approximately 13.5 lakhs of casualties [2]. Due to its specific vehicular composition, Marrakech, a popular city in Morocco, poses a big challenge for traffic flow and safety studies. The traffic in the city is made up of roughly 40% motorcycles and 60% cars, which has a significant impact on traffic dynamics and safety patterns. The mix includes various factors, such as vehicle volumes, road conditions, and traffic control measures. The traffic system in Marrakech has a unique set of challenges because of the high number of motorcycles. In fact, Marrakech's traffic is unique and presents big challenges that have a significant impact on traffic flow and safety [3].

In recent years, there has been a proliferation of studies about traffic safety. Prediction of traffic accidents can help to prevent crashes, avoiding damage, providing drivers with warning alerts to potential risks or identifying potential causes and precautionary measures to prevent crashes from occurring [4]. The purpose of traffic risk studies is to identify crucial factors for future planning scenarios and to examine potential factors that may have adverse effects. According to research, most accidents occur due to driving behavior.

The author in study [5] identified multiple aspects of driver behavior from the same viewpoint and separated two aspects: (1) the research focused on individual psychological determinants as the basis for risky traffic behavior and its predictor, and (2) investigating the specific aspects of traffic climate that relate to social psychological determinants. In addition, they underlined the importance of focusing on studying driver behavior models: (a) The content aspect considers the limits of driver behavior and their distinctions from aggressive or dangerous driving, (b) the structural aspect involves identifying risky road behavior by examining the relationships between various behavioral manifestations of patterns in a wide social context, and (c) the dynamic aspect that identifies the behavior change in future [2].

Through various studies and research in psychology, it has come to light that rewarding desirable behavior can lead to a faster and longer-lasting change in human behavior than punishing undesirable behavior. Research has indicated that rewarding road safety behavior can be helpful, as evidenced using seatbelts and keeping driving speeds within permissible limits. However, a proper reward system is necessary for optimal effect. The authors in the study [6] proposed a system that constantly monitors drivers and uses real-time data to give them an appropriate score, provide feedback accordingly, and help punish errant drivers. They utilized different measured human factors as inputs to their smart algorithm to give an appropriate score to the driver based on their efficiency. Due to that, the major causes of road accidents and global warming are attributed to human factors. In this paper, we proposed an approach which considers human factors that can be analyzed with a smartphone sensor, a car and an OBD-II adapter. The measured human factors are given as inputs to a smart algorithm that delivers an appropriate score to the driver based on his efficiency.

The remainder of the paper is organized as follows: Section II provides an overview of the relevant literature. In Section III, we proposed our methodology in which we presented the data acquisition and data analysis. Section IV compared the different machine learning models and approved the choice of appropriate ML model. Section V covers the experimental and simulation results. Finally, Section VI summarize the paper and discusses future works.

II. Related Work

To evaluate road safety measures based on accident frequency and severity, several researchers have studied the safety effects of different road safety measures [7]. As a result, relevant techniques and methodologies have been developed to detect road accidents and identify critical factors that can cause crashes.

Nericell is a system proposed by study [8] that leverages smartphones to identify various vehicle conditions, such as braking, road bumps, honking, and stop-and-go traffic. It utilizes the smartphone's accelerometer, microphone, GSM communications, and GPS for data collection. The information gathered from multiple smartphones is aggregated on a

centralized server. The detection of such events does not involve the use of machine learning (ML) algorithms. The authors used the pattern matching and orientation calibration to detect the various vehicle situations. Similarly, the mobile sensor-platform for intelligent recognition of aggressive driving (MIROAD) is a driver behavior monitoring system developed by Johnson and Trivedi [9]. This system relies entirely on a smartphone's internal accelerometer, gyroscope, magnetometer, and GPS. They were the first to introduce a more advanced pattern recognition approach. MIROAD-equipped smartphones can detect and classify various aggressive and non-aggressive driving maneuvers using the variations of Dynamic Time Warping (DTW). The DTW is an algorithm to find similar patterns in temporal series. They don't introduce the machine learning models to classify the aggressive and non-aggressive behavior. Likewise, the system proposed by study [10] evaluates a person's driving as either safe or risky. It identifies risky events such as sudden maneuvers, turns, lane departures, braking, and acceleration using only the accelerometer, gyroscope, and magnetometer of a smartphone. It uses an endpoint detection algorithm to identify events and a DTW algorithm to compare input data with template events. Additionally, they incorporate a feature does not present in Johnson and Trivedi's system [9] which is the labeling of driver behavior. A Bayesian classifier labels the driver's behavior as safe or risky based on a calculated probability. Alternatively, the authors in study [11] developed and built a vehicle integrated system to inform drivers about the quality of their driving. They provided a safety index to each driver which indicating their ability to drive safely over prolonged periods. To reduce accidents, the safety index generated can be utilized to reward or retire the drivers. They based on the acceleration, speed and road traffic conditions parameters to indicate the Safety Index. For that, they utilized smartphones and high-end mobile devices having all the sensors to access data to develop an integrated application to track driver's performance. The machine learning models are not involved in their contribution. The authors in study [12] proposed a machine learning model based on set of rules to classify the driving maneuvers from time series data. They categorized the driver behavior as aggressive acceleration, non-aggressive, aggressive right turn, aggressive left turn, aggressive right lane change, and aggressive breaking aggressive left lane change. In fact, they utilized the accelerometer to collect the traffic data during longitudinal, lateral movement and the gyroscope for angular movement.

Recently, authors in study [13] proposed a classification model for drivers' behaviors. The dataset was built from recording OBD II (On Board Diagnostic II) parameters. The recorded data were analyzed and calculated to derived additional driving metrics such as Avg_fuel, Idle_engine, High_speed_breaking, Revv_engine, Dev_str, Vs_dev, Idle_instance, Avg_gear, Hb_instances, Avg_speed, Rev_instances. In this research, they employed the machine learning techniques, like SVM, AdaBoost, and Random Forest to classify driver behavior into ten classes: poorest, poor, bad, belong average, average, good, above average, extremely good, excellent, very good.

In our approach, we classify driving behavior into one of the following classes: aggressive, moderate or slow behavior to overcome the shortcomings of previous research that has focused only on aggressive behavior in various driving events. However, they have not addressed the other cases related to normal or slow behaviors. Additionally, we emphasize that slow driving can have a significant negative impact on urban traffic. A driver moving significantly slower than the flow of traffic may cause frustration for other drivers, leading to sudden lane changes or tailgating. Moreover, to accurately identify moderate and slow driving behaviors, it is essential to incorporate additional data features provided by OBD II adapter such as average speed, acceleration, throttle position, engine revolutions, and the vehicle's geographic location. These data features are crucial for a more comprehensive analysis of driving patterns.

Some significant points have been identified shown in Table I outlined here:

a) All papers use sensor data as input for the event detection algorithms, except [13] which used OBD II to create their dataset.

b) Some of them doesn't utilize the machine learning for classification of driver behaviors.

c) The majority of the research results in analyzing data using either mobile sensors such as accelerometer, gyroscope, magnetometer, GPS or OBD II.

TABLE I. COMPARISON BETWEEN EXISTING DRIVING BEHAVIOR STUDIES

Existing works	Types of Datasets	Machine learning models	Driver behavior Classes	Driver feedback application
[8]	smartphone's accelerometer, microphone, GSM communications, and GPS	Not considered	Not considered	Not developed
[9]	accelerometer, gyroscope, magnetometer, GPS and video	Not considered	non-aggressive and aggressive.	Not developed
[10]	accelerometer, gyroscope and the magnetometer	Bayesian classifier	safe or risky	Not developed
[11]	smartphones, high-end mobile devices and videos	Not considered	Not considered	Not developed
[12]	Accelerometer, gyroscope and the magnetometer	unsupervised learning technique using the sequential covering algorithm and classifying driving maneuvers from time-series data	aggressive acceleration, non-aggressive, aggressive right turn, aggressive left turn, aggressive right lane change, and aggressive breaking aggressive left lane change	Not developed
[13]	OBD II adapter and derived additional driving metrics	SVM, AdaBoost, and Random Forest	poorest, poor, bad, belong average, average, good, above average, extremely good, excellent, very good	Not developed

We deducted that there is no interest in combining smartphone sensors and OBD II features. By combining sensors and OBD II data systems offers more accuracy to classify driving behavior (aggressive, moderate, or slow) and capture the nuances of driver actions. This integrated approach enhances the ability to detect risky driving patterns and improves safety systems by offering a real-time, comprehensive analysis of driving habits.

The purpose of this paper is to present a system having advantage to predict and prevent by warning a driver if his or her driving style is aggressive, moderate or slow and to give information about incidents to gain a better understanding of their causes. Therefore, the proposed system determines the drivers with aggressive labels in localized areas to provide insight into where accidents are more likely to happen.

III. Methodology

Unfortunately, up to now the research involved the driving behavior factor for safety road are not widely studied in the literature. In fact, we proposed a road driving awareness system that focuses on driver behavior as a key factor in reducing road accidents and congestion problems in Morocco, particularly in Marrakesh city. Furthermore, road accidents in Morocco lead to over 4,000 deaths annually and approximately 140 million Dirhams (equivalent to about $14 million) of property damage [14]. During our analysis of the traffic road in Marrakech city, we observed that the behavior of drivers is influenced by various parameters, including infrastructure, vehicle performance, fatigue, and overspeed and we pointed out that motorcycles are the major factor that influences road accidents and congestion. Due to that, we propose a system that can predict various road problems by gathering the driver behaviour in real time. We began our system's development by collecting sensors and OBD II data, analyzing the effects of various features, and using machine learning algorithms to classify the provided data to visualize driver behavior in real-time, which led us to propose an awareness system.

A. Data Acquisition

By integrating data from smartphone sensors and OBD II adapter, driving behavior can be classified more precisely (aggressive, moderate, or slow), allowing for a deeper understanding of driver actions. This combined approach improves the detection of risky driving patterns and enhances safety systems by providing a comprehensive, and real-time analysis of driving habits. Each sensor provides unique data that, when combined, creates a more detailed and accurate picture of a driver's actions. These sensors are:

1) Accelerometer: detects rapid changes in speed, such as hard braking or sudden acceleration, key indicators of aggressive driving behavior (Fig. 1).

2) Gyroscope: measures rotational movement, helping to identify sharp turns or swerving, which can also indicate aggressive or unsafe driving (Fig. 1).

3) Magnetometer: used to determine the vehicle's orientation relative to the earth's magnetic field, aiding in directional awareness.

4) GPS: provides real-time location data, enabling the analysis of driving speed and route choice. It also helps assess whether a driver is adhering to speed limits and can indicate potential traffic conditions.

5) OBD-II: captures vehicle-specific data like throttle position, engine RPM, and fuel consumption, which provide insights into how efficiently the vehicle is being driven.

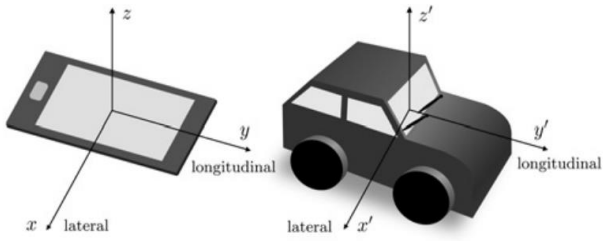


Fig. 1. Accelerometer and gyroscope sensors with vehicle coordinate system.

In our approach, we interest to label the driver behavior in three classes aggressive, moderate and slow behavior. For aggressive behavior, we collected the data set from accelerometer and gyroscope sensors to detect sudden maneuvers and sudden brakes. For moderate behavior, we need some additional features from OBD II like average speed, engine revolutions, throttle to detect average driving maneuvers. In similar way, we classify driver behavior as slow when the driver is maintaining a lower-than-average speed. To detect irregularities in the car, we must connect the OBD functions to a DLC (data link connector). The DLC is a collection of codes that make it easier to detect the work of sensors. The primary function of the OBD system is to monitor the functioning of essential engine components, such as those responsible for regulating emissions and detecting engine defects. Both hardware and firmware are crucial to extracting data from the OBD-II system. The ELM327 Bluetooth-based multi-protocol device located at the core of the system is compatible with the OBD-II protocols as shown in Fig. 2.

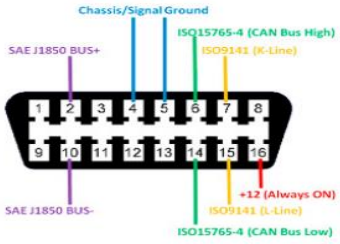


Fig. 2. OBD II system.

B. Data Analysis

To analyze the data, live vehicle data was necessary. It was collected from the Clio 4 vehicle, which was driven by same driver in Marrakesh city way. Fig. 3 illustrates the route utilized for gathering the dataset in a single direction, represented in black colored segment starting from the same point. We selected thoughtfully to simulate urban driving conditions. Therefore, we opted for a route that would enable us to collect the most relevant data possible. It was collected during crucial times, which were

chosen to reflect the different traffic conditions that are commonly observed throughout the day. The peak hours for rush are at 8:00 AM to 8:30 AM, midday from 12:00 PM to 12:30 PM, and evening from 16:30 PM to 17:00 PM. The OBD-II recorded data every second while we were driving and collected 22 features that are average fuel consumption, average speed, average speed (GPS), calculated boost, calculated instant fuel consumption, distance travelled, engine RPM x1000, fuel used, fuel used price, instant fuel power, vehicle acceleration, OBD module voltage, Calculated engine load value, engine coolant temperature, intake manifold absolute pressure, engine RPM, vehicle speed, intake air temperature, throttle position, fuel rail press, speed (GPS), altitude (GPS). At the same time, we collected the longitudinal and lateral movement from accelerometer as acc_x, acc_y, and angular movement from gyroscope gyr_z, dataset using a Samsung Galaxy A31 smartphone which is positioned at the center of a car's windshield. Table II presents the various feature categories along with their descriptions.

Fig. 3. Itinerary chosen for data collection in Marrakesh, Morocco.

TABLE II. THE OBD AND SENSORS FEATURES DESCRIPTION

Category	Features	Description
OBD II	Average speed	The average speed of the vehicle over a period of time, calculated by dividing the total distance traveled by the total time taken
	Average fuel consumption	The average rate of fuel usage, typically measured in liters per 100 kilometers (L/100 km), indicating how efficiently the vehicle uses fuel
	engine RPM x1000	The revolutions per minute of the engine, divided by 1000. It indicates how fast the engine's crankshaft is rotating.
	Distance travelled	The total distance covered by the vehicle during a trip, typically measured in kilometers
	Average speed (GPS)	The average speed of the vehicle as determined by the GPS system. It may differ from the calculated average speed due to factors like GPS signal accuracy
	vehicle acceleration	The rate at which the vehicle's speed increases, typically measured in meters per second squared (m/s²)
	OBD module voltage	The voltage level of the OBD module, typically corresponding to the vehicle's battery or alternator voltage
	Calculated engine load value	The calculated value representing the percentage of the engine's maximum capability being used. It provides an idea of how hard the engine is working
	Engine coolant temperature	The temperature of the engine's coolant, indicating how hot the engine is running.
	Intake manifold absolute pressure	The pressure inside the intake manifold, typically measured in kilopascals (kPa) or bar, used to calculate engine load and airflow
	Engine RPM	The number of revolutions per minute (RPM) at which the engine's crankshaft rotates, indicating engine speed
	Vehicle speed	The current speed of the vehicle, usually measured in kilometers per hour (km/h)
	Intake air temperature	The temperature of the air entering the engine, which can affect performance and fuel efficiency
	Throttle position	The position of the throttle valve, typically measured as a percentage from fully closed (0%) to fully open (100%), indicating how much air is being allowed into the engine
	Fuel rail press	The pressure of the fuel in the fuel rail, typically measured in kilopascals (kPa), indicating how much fuel pressure is being supplied to the injectors
	Speed (GPS)	The real-time speed of the vehicle as measured by GPS, typically more accurate than the speedometer reading
	Altitude (GPS)	The vehicle's altitude above sea level as measured by GPS, typically in meters
Accelerometer	acc_x	It measures the rate of change in velocity along the **x-axis**. This typically corresponds to motion in the left-right direction
	acc_y	It measures the acceleration along the **y-axis**, capturing vertical motion (up and down). It detects forces like gravity or upward/downward movement in space
	acc_z	It measures the acceleration along the **z-axis**, typically in the forward-backward direction. This axis is perpendicular to both the x and y axes, capturing motion like moving a device toward or away from a surface
Gyroscope	gyr_x	It measures the angular velocity around the **x-axis**. This represents the rate of rotation around the horizontal axis that runs from left to right
	gyr_y	It measures the angular velocity around the **y-axis**. This axis runs from front to back, so the gyroscope detects rotations in the left or right direction (side-to-side tilting)
	gyr_z	Measures the angular velocity around the **z-axis**. The z-axis runs vertically through the device (up-down direction), it detects rotational movements like spinning or turning the device around this vertical axis

Both datasets were merged by data acquisition module developed on mobile application and sent in real time for storing in NoSQL database. In this work, we assume that a driving behavior can be labeled into one of the classes described in the following subsections.

a) Aggressive behavior: As previously stated, we keep a gyroscope and accelerometer to recognize risky behavior. The accelerometer and gyroscope value are utilized during the trip to detect driving events like hard braking, sudden acceleration, or aggressive turning. To ensure non aggressive driving, it is recommended to use the recommended range of acceleration or deceleration values presented by [11][12][15]. For safe accelerating and braking the acc_x should be between -3 to 3 m/s². In contrast, the maneuvers such as left turn (LT), right turn (RT) influence the lateral and angular changes in a motion. Over the duration of safe LT and RT, acc_y between -1.5 to 1.5 m/s². Similarly, gyr_z perform between -0.4 to 0.4 m/s². In fact, we can identify aggressive driving, typically, when the threshold for aggressive acceleration rate greater than 2.5 to 3.0 m/s². This means a sudden increase in speed, indicating rapid acceleration. The threshold for aggressive braking rate (negative acceleration) greater than -2.5 to -3.0 m/s². In case of aggressive corning, the lateral acceleration can be greater than 1.5 m/s² and a yaw rate exceeding 15 to 20 degrees per second can indicate sharp turns. We summarized the different type of aggressiveness behavior in Table III.

TABLE III. CATEGORY OF AGGRESSIVENESS BEHAVIOR SEVERITY

Degree of Behavior aggressiveness	Type of Behavior aggressiveness
acc_x greater than -2.5 to -3.0 m/s	Braking with high speed
acc_x 2.5 to 3.0 m/s².	Sudden increase in speed
acc_y greater than 1.5 m/s and gyr_z exceed 15 to 20 degrees	Zig-zag with high speed

b) Moderate behavior: The proposed method considers a good average speed between 50 and 60 km/h. In addition, the optimal engine load indicates how much power the engine is producing relative to its maximum capacity. The normal driving rate its between 20% to 60% of maximum engine load.

c) Slow behavior: Driving the vehicle at a speed that is significantly less than the posted limits or traffic flow. Furthermore, being too cautious can result slow acceleration, early braking, and hesitancy at intersections which could lead to delays and hinder the flow of traffic.

Our research does not assume that driver behavior will remain consistent during the entire trip. Instead, we aim to detect varying actions taken by drivers during the same trip and classify them. In data analysis process, we used gradient boosting on decision trees (GBDT) models as XGBoost, Adaboost, Gradient boosting and decision tree to identify the three potential driving behavior classes. The GBDT algorithms gathers multiple weak learners, typically a decision tree, to create a stronger model that surpasses the base model. When boosting decision trees are added, they learn from the errors of previous individual trees. Each tree is connected in a series and tries to reduce the error of the previous tree. Boost algorithms may be slow to learn because of this sequential connection, but they are great at performing in the classification. In addition, we utilized random forest model which uses bagging techniques to create multiple decision trees using bootstrapped samples. The bagging method produces random samples with replacements from the input data and instructs decision trees based on the samples. A Random Forest classifier trains multiple decision trees on the same training set to enhance classification precision and combat overfitting [13][16].

IV. APPROPRIATE MACHINE LEARNING MODEL

One of the key steps is data preprocessing, which includes data clean-up and normalization. Then, we employed machine learning models for classification. We adopted in this paper XGBoost, AdaBoost, gradient boosting, decision trees, and random forests, to identify three potential driving behavior classes: aggressive, moderate, and slow. The classification techniques utilized here are outlined in the following subsections.

a) XGBoost model: XGBoost is increasingly used in the field of driver behavior analysis due to its effectiveness in handling complex, structured datasets and its ability to provide high predictive accuracy. It's a supervised algorithm that combines multiple base learners to enhance the performance of strong learners.

To analyze driver behavior, we train the model on sensors and OBD datasets $D_{sensors} = \{(X_i, y_i)\}$, $D_{obd} = \{(X_j, y_j)\}$ containing n samples and m decision variables. The driver behavior class y_i is predicted by the trained model. y_i is the target value corresponding to X_i . To understand the functions within the model, the learning objective function of XGBoost is defined as follows:

$$= \sum_{i=1}^{n} l(\hat{y}_i, y_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (1)$$

where, l is a differentiable convex function that measures the discrepancy between the predicted classification value $\hat{y}_i$and the actual category y_i, this is known as the loss function. The Ω function represents the regularization term, which penalizes model complexity to help prevent overfitting.

b) AdaBoost model: AdaBoost algorithm, a widely used ensemble classifier. It works by sequentially training multiple base learners on different sets of training data. The base learners are generated sequentially. The first learner, L_1 is trained on a random subset of the data. The second learner, L_2 is trained using data where half are correctly classified by L_1 and the other half are misclassified by L_1 . The third learner, L_2 focuses on examples misclassified by both L_1 and L_2. The algorithm assigns a weight, w_1, to each training sample. In each iteration, these weights are updated, particularly increasing the weights for the misclassified samples, to help the learner focus on harder cases. The final classifier chosen in each iteration, h_r , is assigned an importance score, α_r , based on its mean squared error e_r, according to the formula:

$$\alpha_r = \frac{1}{2} \ln\left(\frac{1-e_r}{e_r}\right) \quad (2)$$

This process continues iteratively, and different base learners work together to reduce the classification error.

c) Gradient boosting model: Gradient Boosting Machine (GBM) is one of the most widely used supervised machine learning techniques. It was introduced by [17] and has been shown to be highly effective for various predictive tasks. While the training process of GBM, a function is used to distinguish between aggressive, moderate and slow driving behavior. The dataset contains O observations, each characterized by features. The training set is represented as X_i, where related labels y_i with $y_i \in \{0,1,2\}$ where 0 indicates an aggressive behavior, 1 indicates a moderate behavior and 2 indicates a slow behavior.

The goal is to minimize the aggregated loss function through multiclass function estimation. This is typically expressed as an extension of binary classification, where multiple functions are estimated, one for each class. Instead of estimating a single function, the model estimates a separate function for each class label. This is done using the following formula:

$$P(y = N|X_i) = \frac{\exp(f_N(X_i))}{\sum_{j=1}^{N} \exp(f_j(X_i))} \quad (3)$$

Where $f_N(X_i)$ is the function learned for class N at input X_i , $P(y = N|X_i)$ is the predicted probability that the instance belongs to class N, exp denotes the exponential function.

d) Random forest model: Random forest (RF) model utilizes a method called bootstrap aggregating (also known as bagging) to create multiple decision trees and enhance model performance. Bagging involves selecting random subsets of features to determine the best split points in the tree-growing process. For classification tasks, RF bases the final prediction on a majority vote across all trees. The RF predictor function is represented as:

$$f(x) = \frac{1}{K} \sum_{i=1}^{K} T(x) \quad (4)$$

where K is the number of trees, and T(x) represents the individual tree predictions. RF reduces the correlation between trees by using bagging, which creates diversity among trees by drawing different data subsets with replacement. This resampling method allows some data points to be reused multiple times, while others might not be used at all. The diversity enhances robustness to slight variations in input data.

When growing trees, RF selects the best split point from a randomly chosen subset of features at each node. This decreases the strength of individual trees but reduces their correlation, ultimately improving generalization.

e) Decision trees model: The decision tree algorithm is trained using labeled data where each driving behavior is tagged. The tree is constructed by recursively splitting the data based on the most informative features. At each node of the tree, a specific feature is used to split the data, resulting in child nodes representing subsets of the data with more homogeneous driving behaviors. This process continues until the data is sufficiently categorized, or a stopping criterion (such as maximum depth or minimum samples) is met. At each node, the best feature $X_i$and threshold θ are selected to maximize the information gain or minimize Gini impurity. The Gini impurity formula used to assess the purity of a node is:

$$Gini(T) = 1 - \sum_{i=1}^{N} p_i^2 \quad (5)$$

Where N is the number of classes (e.g., "aggressive", "moderate", "slow") and p_i is the proportion of data points in class i at node T.

To assess the efficiency of driver behavior classification, we compare the performances of XGBoost, AdaBoost, Gradient boosting, Decision tree, Random Forest using the following metrics:

- Macro Accuracy: accuracy is the ratio of correctly predicted instances (both positive and negative) over the total number of instances. In multi-class problems, macro accuracy is the overall accuracy across all classes, treating each class equally.

$$\text{Average Accuracy} = \frac{1}{N} \sum_{i=0}^{N} \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \quad (6)$$

- Macro Precision: precision measures how many of the predicted positive instances were true positive. It's the ratio of true positives (TP) to the sum of true positives and false positives (FP). For multi-class classification, macro precision is the unweighted average precision across all classes.

$$\text{Macro Precision} = \frac{1}{N} \sum_{i=0}^{N} \frac{TP_i}{TP_i + FP_i} \quad (7)$$

- Macro recall is used to evaluate the performance of a classification model, particularly in multiclass problems. It measures the model's ability to correctly identify positive instances for each class and then averages these recalls without considering class sizes.

$$\text{Macro Recall} = \sum_{i=0}^{N} \text{Recall}_i = \frac{1}{N} \sum_{i=0}^{N} \frac{TP_i}{TP_i + FN_i} \quad (8)$$

- The macro F1-score combines the concepts of precision and recall into a single score, providing a balanced measure of a model's performance across all classes.

$$\text{Macro F1} - \text{score} = \frac{1}{N} \sum_{i=0}^{N} \frac{\text{Precision}_i \, \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \quad (9)$$

Where N is behavior classes number which means the three classes in our study, and TP_i, FP_i, TN_i, and FN_i are the true positives, false positives, true negatives, and false negatives. Macro-averaging approach is employed to solve the problem of imbalanced data when calculating accuracy, recall, precision, and F1-score.

We divided our dataset into 70% of training and 30% of testing. Tables IV and V presented the comparison of different machine learnings algorithms based on sensors and OBD II data, respectively, using the metrics described above. We deduct that Random Forest model has achieved an accuracy of 96% and 99% in both sensors and OBD testing sets. The Random Forest

model demonstrated his efficiency in multi classification tasks. In fact, we chose it as the appropriate machine learning model to predict the three driver behavior classes already proposed.

TABLE IV. PERFORMANCE OF THE DIFFERENT MACHINE LEARNINGS MODELS USING SENSORS DATABASE

Metrics	ML models				
	XGBoost	*AdaBoost*	*Gradient Boosting*	*Decision Tree*	*Random forest*
Average Accuracy	92%	87%	88%	90%	96%
Macro Precision	91%	72%	90%	79%	97%
Macro Recall	72%	52%	54%	78%	82%
Macro F1-score	78%	51%	54%	79%	88%

TABLE V. PERFORMANCE OF THE DIFFERENT MACHINE LEARNINGS MODELS USING OBD DATABASE

Metrics	ML models				
	XGBoost	*AdaBoost*	*Gradient Boosting*	*Decision Tree*	*Random forest*
Average Accuracy	99%	96%	97%	98%	99%
Macro Precision	95%	87%	91%	80%	98%
Macro Recall	82%	84%	70%	77%	86%
Macro F1-score	88%	85%	79%	78%	92%

V. SIMULATION

Fig. 4 illustrates an overall architecture for visualizing driver behavior from data acquisition to classification and driver behavior prediction in real-time such as aggressive, moderate or slow. This architecture is based on three principal keys: (1) Sensing required data from driver's environment and context, (2) Analysing and labeling the driver's behavior class based on the gathered data, and (3) Visualizing the obtained label on driver's mobile application. These main steps are presented in more detail in the following subsections.

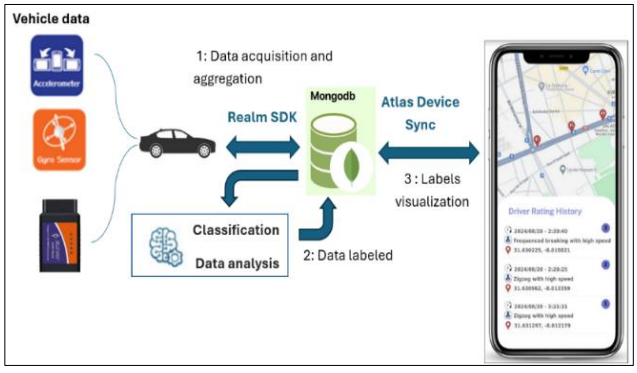


Fig. 4. Smart system for driver behavior prediction.

C. Data Acquisition and Aggregation

As mentioned earlier, the driver behavior prediction objective needs to collect various features that can help the system to categorize driver behavior into three classes: aggressive, moderate, and slow. These features are gathered both from the driver's mobile phone and from an OBDII that is connected to the driver's car and his mobile phone. The collected data related these features are concatenated in a single row json to be stored at real-time in a NoSql Mongodb database. Then the inserted data will be analysed by the chosen ML model which is Random Forest. To meet the challenge of communicating at real-time these data to Mongodb database, we used the following technologies:

- MongoDB Realm which is a lightweight database,
- Realm SDKs to synchronise in real-time these data,
- MongoDB Atlas that provides cloud-hosted managed instances of MongoDB for availability.

MongoDB Realm enables synchronisation between Realm Database and MongoDB Atlas, seamlessly stitching together the two components into an application layer for mobile application. This advantage comes from the fact that the core of Atlas MongoDB service is based on call rest API and push services.

D. Data Labeled

Once the data is stored in MongoDB database, Atlas trigger fires the module "Classification Data Analysis" implementing random forest algorithm to predict the corresponding driver behavior. The obtained label will then be updated in Mongodb database.

The random forest model read the row recently inserted in Mongodb. It analyses first the value that comes from mobile sensors to predict if the driver's behavior was aggressive or not. If the detected behaviour is aggressive, the algorithm is considered finished. Otherwise, the algorithm starts the second process by reading the OBD data to predict if the driver's behavior is moderate or slow.

Once the class is determined by the random forest model, the obtained label is then updated in MongoDB database.

The detail of the implemented algorithm is presented as follows:

Algorithm 1: Prediction Driver behavior

1. **Input**: Each Last added Row from MongoDB at real-time
2. {dataOBD, dataSensors}
3. dataOBD: Data related to OBD columns
4. dataSensors: Data related to Sensor's columns
5. **Output**: The variable "Behavior": {Agg, Slow, Moderate}
6. **Begin**
7. # **First step** – Prediction of aggressive behavior case
8. Behavior = Prediction of the behavior related to dataSensors
9. # **Second step** – Prediction of Moderate or Slow behavior cases
10. If (Behavior! = Agg)
11. Behavior = Prediction of the behavior related to dataOBD
12. # Return the result Aggressive, Moderate or Slow
13. Return Behavior
14. **End**

E. Labels Visualization

The objective of this application is not only to provide driver with valuable feedback about their driver behavior but in addition, to motivate them to be more attentive and conscious of their actions. The application was built using Flutter technology, a well-known open-source framework for designing robust and multi-platform mobile applications. Users can receive real-time updates and access to the most up-to-date information thanks to its seamless integration with APIs. Flutter technology is connected to Atlas Device Sync in order get the updated rows in MongoDB done by the classification ML model module (step 2). The map is used to display pertinent information about driver behavior in real-time using machine learning algorithms, and each color represents a different type of driver behavior. An example of driver interface is shown in the Fig. 5.

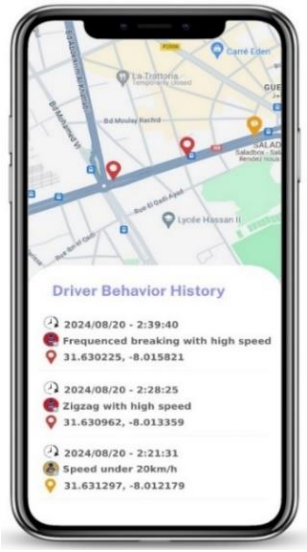


Fig. 5. Driver behavior history application.

VI. Conclusion and Future Work

Driving behavior is a crucial problem that affect the driver safety. To overcome the problem, we proposed a mobile application to aware the driver about his/her driving behavior styles. In this contribution, we developed a smart system to predict driver behavior in real-time utilizing machine learning model and the data collected from smartphone sensors and OBD II installed in a car. This system has the advantage to display on the map the relevant information about driver behavior predicted by machine learning algorithm in real time by assigning different color in which identifies a different type of driver behavior. Simulation have shown the ability of random forest model to predict three classes: aggressive, slow, moderate in both OBD and sensors data to achieve an accuracy of 92% and 88% respectively in Marrakesh city as case study. As a future work, our upcoming project involves developing a real-time recommendation system for safer routes based on driver behaviors. To establish safe routes through intersections, we propose an approach that considers the safety index of each road segment. This index will be calculated in real time, based on the safety indices of vehicles currently traveling on the respective road segment at any given moment. The safety index for each vehicle at time t will be computed in real time using the aggressiveness classification model (which distinguishes between aggressive and slow behaviors) and subsequently stored in a MongoDB repository.

References

[1] WHO/T. Pietrasik. 2020 https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

[2] H. Lazar, Z. Jarir, "Road traffic accident prediction: a driving behavior approach". 8th International Conference on Optimization and Applications (ICOA), 1–4. 2022. doi: 10.1109/ICOA55659.2022.9934000

[3] A. Charef, Z. Jarir, and M. Quafafou. "The Impact of Motorcycle Positioning on Start-Up Lost Time: The Empirical Case Study of Signalized Intersections in Marrakech using VISSIM". Eng. Technol. Appl. Sci. Res., vol. 14, no. 3, pp. 14313–14318, Jun. 2024.

[4] I. J. Mrema and M. A. Dida, "A Survey of Road Accident Reporting and Driver's Behavior Awareness Systems: The Case of Tanzania", Eng. Technol. Appl. Sci. Res., vol. 10, no. 4, pp. 6009–6015, Aug. 2020.

[5] TV. Kochetova. "The Patterns of Drivers' Traffic Behavior: Evidence From Three Countries". Front Psychol. 2022 Apr 7;13:869029. doi: 10.3389/fpsyg.2022.869029. PMID: 35465507; PMCID: PMC9021888.

[6] S Khedkar, A Oswal, M Setty, S Ravi. "Driver Evaluation System Using Mobile Phone and OBD-II System". Int. J. Comput. Sci. Inf. Technol. Vol. 6 (3) , 2015.

[7] G. Yannis, A. Dragomanovits, A. Laiou, F. La Torre, L. Domenichini, T. Richter, S. Ruhl, D. Graham, and N. Karathodoro. "Road traffic accident prediction modelling: a literature review." Proceedings of the Institution of Civil Engineers - Transport 2017 170:5, 245-254 http://dx.doi.org/10.1680/jtran.16.00067

[8] P.Mohan, V.N.Padmanabhan, R.Ramjee,: "Nericell: rich monitoring of road and traffic conditions using mobile smartphones". Proc. of the Sixth ACM Conf. on Embedded Network Sensor Systems, 2008, pp. 323–336.

[9] D.A., Johnson, M.M.Trivedi. "Driving style recognition using a smartphone as a sensor platform". 14th Int. Conf. on Intelligent Transportation Systems (ITSC), 2011, pp. 1609–1615.

[10] H.Eren, S. Makinist, E.Akin, A.Yilmaz. "Estimating driving behavior by a smartphone'. Intelligent Vehicles Symp. (IV), 2012, pp. 234–239.

[11] S. Chigurupati, S. Polavarapu et al., "Integrated computing system for measuring driver safety index," Int. J. Emerg. Technol. Advanced Eng., vol. 2, no. 6, 2012.

[12] M. M.Haque, S. Sarker, M.A.A. Dewan. "Driving maneuver classification from time series data: a rule based machine learning approach". Appl Intell 52, 16900–16915. 2022. https://doi.org/10.1007/s10489-022-03328-3

[13] R.Kumar, A. Jain. "Driving behavior analysis and classification by vehicle OBD data using machine learning". The Journal of Supercomputing. 2023 Nov;79(16):18800-19.

[14] M. Benhadou, I. Chairi, and A. Lyhyaoui." Accident Prediction Model - Case Study Region of Tangier". Journal of Traffic and Logistics Engineering, 23-27. 2019.

[15] J. Ferreira, E. Carvalho, B, V. Ferreira, C. de Souza and Y. Suhara, A. Pentland, and G. Pessin.,"Driver behavior profiling: An investigation with different smartphone sensors and machine learning, PLoS one, 12(4), 2017.

[16] M.Domor, S.Yanxia.. A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. IEEE Access. PP. 1-1.2022. 10.1109/ACCESS.2022.3207287.

[17] Jerome H..Friedman."Greedy function approximation: A gradient boosting machine".Annals of Statistics 29 (2001): 1189-1232.

Development of Intelligent Learning Model Based on Ant Colony Optimization Algorithm

Xiaojing Guo[1]*, Xiaoying Zhu[2], Lei Liu[3]
College of Technology and Data, Yantai Nanshan University, Yantai 265700, China

***Abstract*—In the process of the gradual popularization of online courses, learners are increasingly dissatisfied with the recommendation mechanism of imprecise courses when faced with a large number of course choices. How to better recommend relevant courses to targeted users has become a current research hotspot. An intelligent learning model based on ant colony optimization algorithm is introduced, which can accurately calculate the similarity between courses and learners. After structured classification, the model recommends courses to learners in the optimal way. The results showed that the accuracy of this method reached 10-20 when tested in Sphere and Ellipse functions, and the optimal solution for problem Ulysses21 was 27, which was better than Advanced Sorting Ant System (ASrank), Maximum Minimum Ant System (MMAS), and Ant System (AS) based on optimization sorting. The proposed ant colony optimization algorithm had better convergence performance than ASrank, MMAS, and AS algorithms, with a shortest path of 53.5. After reaching Root Mean Square Error (RMSE) and Relative Deviation (RD) distributions of 6% and 8%, the stability of the proposed method no longer decreased with increasing RMSE. The accuracy did not vary significantly with changes in the dataset, and the reproducibility performance was better than other comparison models. In the scenarios of path Block and path Naive, the proposed algorithm had an average computation time of only 1011, which was better than the Ant Colony Optimization (ACO) and Massive Multilingual Speech (MMS) models. Therefore, the proposed algorithm improves the performance of intelligent learning models, solves the problem of local optima while enhancing the convergence efficiency of the model, and provides new solutions and directions for increasing the recommendation performance of online learning platforms.**

***Keywords*—Online courses; ant colony optimization algorithm; intelligent learning model; path planning; local optimum**

I. Introduction

With the development of the Internet, online courses have become more and more popular, which has greatly promoted the development of education in China [1-2]. The booming development of online courses has led to a rapid increase in the number of online courses. However, with the explosive growth of online courses, students are facing unprecedented information overload problems. Traditional recommendation systems often rely on simple user behavior or content features for course matching, making it difficult to accurately associate personalized student needs with corresponding courses, resulting in the dilemma of "rich information but difficult selection" [3-4]. The mismatch between students and courses not only reduces students' learning efficiency, but also restricts the further improvement of the service quality of online education platforms [5]. Therefore, there is an urgent need for an intelligent course recommendation system to solve the above problems.

To address the above challenges, more and more experts and scholars are combining machine learning with path planning, aiming to achieve more efficient personalized recommendation strategies. Liu Y and other researchers proposed a Levy-based ACO algorithm. This algorithm utilized the Levy flight model to expand its search range while searching for paths. The results indicated that its search performance was superior to existing travel agent path planning algorithms [6]. Gao W and other researchers proposed a new ACO algorithm. This algorithm utilized the combination search function to solve the problem of spatial complexity. The results showed that the algorithm outperformed existing common ant colony algorithms in terms of convergence speed and search efficiency [7]. Liu Y and other researchers proposed an ant colony algorithm based on greed and Levy flight improvement. This algorithm adopted pseudo randomness and balanced speed and space to solve the local optimal problem of path search. The results showed that the algorithm performed better than other algorithms when applied in travel sales scenarios [8]. Stodola P and other scholars proposed a car path planning algorithm based on ant colony algorithm. This algorithm used the Cordreau benchmark instance to solve the problems of other methods. The results showed that this algorithm had higher accuracy and precision than other related algorithms [9]. Zhang G et al. proposed an ACO algorithm based on genetic variation, which introduced methods such as crossover, recombination, and mutation in genetics to solve the problem of ship meteorological route planning. The results indicated that this algorithm had higher accuracy and applicability than other algorithms [10].

In recent years, many scholars have devoted themselves to building intelligent learning models to improve the existing performance of algorithms in data similarity calculation research. Intended to enhance the ability of intelligent learning models in processing large-scale complex data, improving classification accuracy, and accelerating computational efficiency. Li W and other researchers proposed a learning classification method based on intelligent optimization algorithms. This method combined operators to assign learning mechanisms to specific learning abilities, achieving analysis and classification of complex learning scenarios. As a result, this algorithm achieved better classification accuracy and reliability than other algorithms [11]. Ge Q and other scholars proposed an intelligent learning model for path

*Corresponding Author

aggregation through an efficient single source VecSim algorithm. This model used a threshold filtering algorithm to remove items with low similarity using a set threshold, and used a sampling algorithm to estimate the probability of encountering additional paths. The results indicated that the proposed method could effectively accelerate the calculation speed, with a query speed of 0.1 seconds and an error of only 10^{-4} [12]. Aggarwal K et al. developed an intelligent learning model based on artificial intelligence, machine learning, and deep learning. This model could effectively identify, analyze, and made decisions when facing large amounts of complex data, increasing the automation of data processing. The results indicated that the algorithm achieved good data analysis performance in the healthcare field [13]. Janiesch C et al. proposed an intelligent machine learning model based on neural networks, which could effectively distinguish relevant concepts in the fields of electronics and network business. The results indicated that the model outperformed traditional data analysis methods and shallow machine learning models in terms of performance. Yang K et al. proposed a federated machine learning model based on aerial computing. This model utilized multiple access channels combined with stacked waveforms to overcome the bottleneck of machine learning model aggregation. The results indicated that the model increased signal propagation strength and reduced model aggregation error [14]. Mousavinasab E et al. proposed a machine learning model based on action condition rule reasoning, which could redefine learners, classify and aggregate them. The results indicated that this method could achieve personalized learning recommendations and promote the application of learning in physics, chemistry, and clinical fields [15].

The existing models still face many challenges in dealing with the complex user behavior patterns and course features unique to online education platforms, such as the dynamic changes in user interests, the diversity of course content, and the real-time nature of recommendation results. Therefore, the study introduces the Negative Feedback Ant Colony Algorithm (NFACA) and the construction of intelligent learning models to be applied to the platform course recommendation platform, helping to improve the performance of platform recommendation algorithms. The innovation of this research lies in the development of an intelligent learning model based on NFACA mechanism, which improves the efficiency and accuracy of path planning and provides learners with more targeted course recommendations. The research contribution lies in proposing an intelligent recommendation model based on NFACA, which provides a new idea and method for course recommendation on online education platforms, solves the problems of low accuracy and efficiency of traditional recommendation algorithms, and provides reference and inspiration for future recommendation system development.

II. Methods and Materials

Firstly, an NFACA is proposed to explore the updating effects of the worst and best solutions in the negative feedback mechanism on pheromone concentration. After defining the learner, a course classification model is constructed based on the calculation method of pheromone concentration, and the intelligent learning model is optimized to improve its performance and find the optimal solution in a shorter time.

A. Design of NFACA Mechanism

Due to the rapid development of society, machine intelligence learning-related technologies have also been further improved. Introducing path planning to find the optimal solution on online learning platforms will become a new driving force for promoting educational progress. The ACO algorithm is inspired by the foraging process of ants, where food always appears randomly. Therefore, the way it selects the optimal path when searching for food has reference significance for model construction in various fields [16]. The use of the ant colony algorithm for path planning mainly solves two problems: optimal path selection and pheromone updating [17]. The selection of the optimal path involves the probability of choosing the next location from one location, as shown in Eq. (1).

$$P_{ij}^{k}(t)=\begin{cases}\dfrac{[\tau_{ij}(t)]^{\alpha}\times[\eta_{ij}(t)]^{\beta}}{\sum_{k\in allowed_k}[\tau_{ij}(t)]^{\alpha}\times[\eta_{ij}(t)]^{\beta}}, \text{if jeallowed}_k\\ 0, \text{otherwise}\end{cases} \tag{1}$$

In Eq. (1), represents the concentration of pheromones; t is the time for selecting the next location; i is the current position; j is the next selected location $\tau_{ij}(t)$ at this moment; $\eta_{ij}(t)$ is the visible range of the ant's next location selection; represents the weight of information concentration; β represents f α actors that affect information concentration; allowed_k is the total number of locations that the k th ant has not yet reached; $P_{ij}^{k}(t)$ represents the probability of ants choosing the next location at this moment. After passing through ants multiple times on a certain road, the concentration of pheromones updates as shown in Eq. (2).

$$\tau_{ij}(t+1)=(1-\rho)\times\tau_{ij}(t)+\sum_{k=1}^{m}\Delta\tau_{ij}^{k}(t) \tag{2}$$

In Eq. (2), $\tau_{ij}(t+1)$ is the pheromone concentration of the ij channel at time $\tau_{ij}(t+1)$; $\Delta\tau_{ij}^{k}(t)$ represents the concentration of pheromones left by ant k on the ij road; m represents the total number of ants passing through ij road; ρ represents the concentration of effective information. The overall process of ant colony algorithm is shown in Fig. 1 [18].

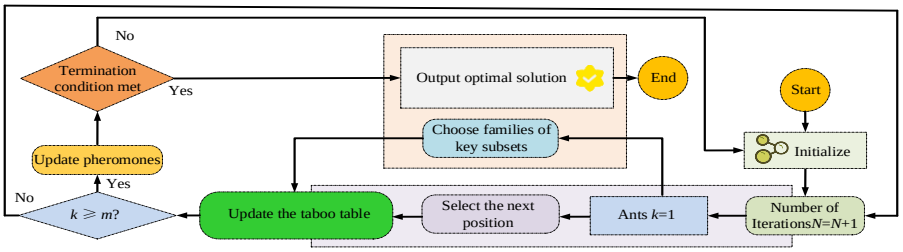


Fig. 1. General flow of ant colony algorithm.

The basic ant colony algorithm takes all the pheromones released by ants as references, which can lead to the algorithm quickly getting stuck in local optima. Therefore, the pheromones released by ants are weighted to avoid their limited field of view and see more possible paths. The weighting formula is shown in Eq. (3).

$$\tau_{ij}(t)=(1-\rho)\tau_{ij}(t)+\sum_{k=1}^{\omega-1}(\omega-k)\Delta\tau_k(i,j)+\omega\Delta\tau_b(i,j) \tag{3}$$

In Eq. (3), w represents the number of ants selected to release pheromones based on their path planning distance sorting; $\Delta\tau_b(i,j)$ represents the concentration of pheromones released by ants that have planned the optimal path; $\Delta\tau_k(i,j)$ represents the concentration of pheromones released by the total number of ants that passed through before k. Only ants that find the optimal path are selected to release pheromones, which can avoid the algorithm's "premature" mechanism in selecting the optimal path. The pheromone optimization formula is shown in Eq. (4).

$$\tau_{ij}(t)=(1-\rho)\tau_{ij}(t)+\Delta\tau_{ij}^{best} \tag{4}$$

In Eq. (4), $\Delta\tau_{ij}^{best}$ is the total concentration of pheromones on the optimal path chosen by ants. The above introduces thinking about optimal path planning from the perspective of the optimal solution, and there is also a reverse thinking entry point: the worst solution, also known as negative feedback. Negative feedback can be used to avoid falling back into the worst solution next time, just like guiding ants to the optimal solution. The formula for the negative feedback model is shown in Eq. (5).

$$\mathbf{Y}=[\delta-\varphi_{ij}(t)]^{\gamma} \tag{5}$$

In Eq. (5), δ represents the upper limit of negative feedback pheromones released on the worst planned path; $\varphi_{ij}(t)$ represents the worst path pheromone matrix; γ represents the weight of negative feedback pheromones on the path; Y represents the probability of negative feedback. The calculation steps of the negative feedback algorithm are shown in Fig. 2.

In the process of combining the worst and best solutions for path planning, different pheromones are used to select paths, and trajectories that are not within the optimal and worst ranges are updated, as shown in Eq. (6) [19].

$$\xi=\frac{L_{\text{best}}}{L_{\text{better}}} \tag{6}$$

In Eq. (6), L_{best} represents the optimal path set that conforms to the concentration of pheromones; L_{better} represents the optimal path set that conforms to the information concentration; ξ represents the relationship and similarity between the optimal and optimal paths. The paths that are not within the optimal and better range are updated, as shown in Eq. (7).

$$\chi_{ij}(t+1)=(1-\rho(t))\times\chi_{ij}(t)+\Delta\chi_{ij} \tag{7}$$

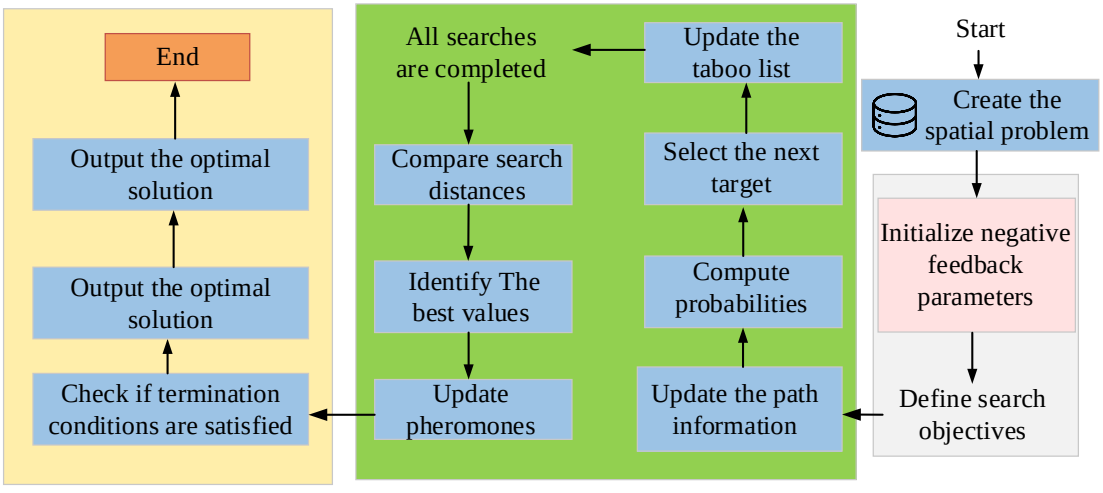


Fig. 2. Computational steps of the negative feedback algorithm.

In Eq. (7), $\chi_{ij}(t+1)$ is the total penalty amount on the channel during $t+1$ time; $\chi_{ij}(t)$ is the total amount of punishment on the ij channel within time t; $\Delta\chi_{ij}$ is the total amount of punishment that varies per unit time; $\rho(t)$ is the amount of remaining pheromone loss within time t [20].

B. Development of Intelligent Learning Path Recommendation Model Based on ACO Algorithm

The primary issue to be addressed in the construction of intelligent learning path recommendation models is the psychological burden that learners face when dealing with a large number of courses. Because a large number of unsorted or unclassified courses are directly pushed without being classified and organized by algorithms, it can cause difficulties for learners to make choices. To structurally classify massive courses and push them along the optimal path, the first step is to classify learners, which can be achieved through accurate descriptions of their learning status [21]. The description of learning state is defined based on the learner's mastery of the course, as shown in Fig. 3.

As can be seen from Fig. 3, the learner's state is divided into start, adaption, and adjustment according to its definition. After defining learners, it is necessary to clarify the calculation of pheromone concentration in ant colony algorithm, which is of great significance for the probability of selecting which path in the path recommendation process. The concentration of pheromones is calculated based on the learner's historical learning curriculum set. The concentration of pheromones on this path not only needs to divide learners into corresponding structural models based on their learning status, but also needs to divide the course into modules corresponding to learners. Therefore, the intelligent learning model requires three datasets for course recommendation: learner dataset, learning course dataset, and pheromone concentration database. The recommendation process is shown in Fig. 4.

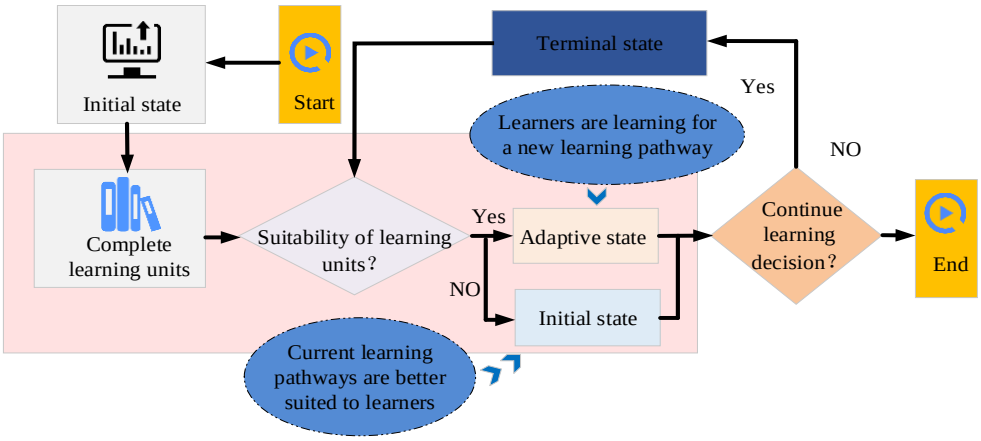


Fig. 3. The transfer diagram of learning states.

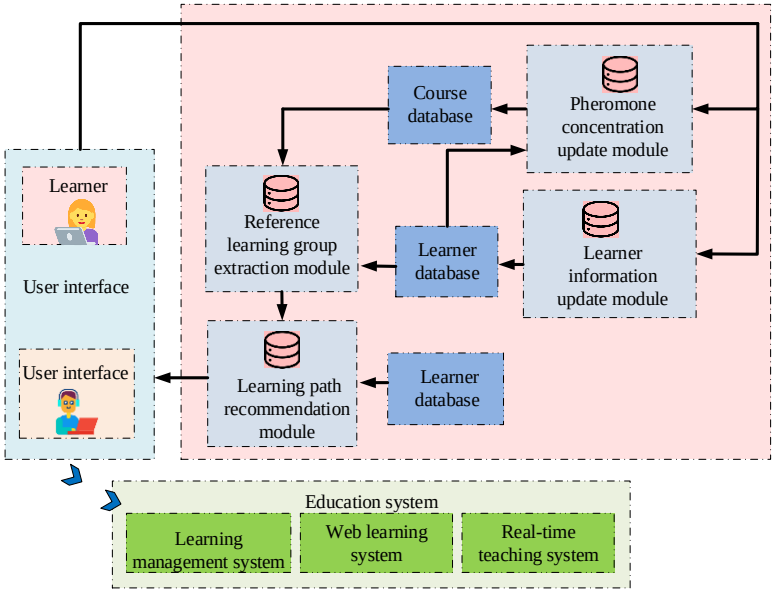


Fig. 4. Study program recommended path.

Learner data information needs to be kept updated, which is the basis for selecting the latest path based on positive and negative feedback. This ensures that each recommended path is not dependent on historical judgments and can be adaptively adjusted based on data updates. This data update requires continuous similarity calculation of learner information, as shown in Eq. (8).

$$\mathrm{Dis}\left(\mathrm{LA_i},\mathrm{LA_j}\right)=\mathrm{Dis}\left(\mathrm{LD_i},\mathrm{LD_j}\right)\times\mathrm{Dis}\left(\mathrm{E_i},\mathrm{E_j}\right) \tag{8}$$

In Eq. (8), $\mathrm{LA_i}$ and $\mathrm{LA_j}$ respectively represent two different learners; $\mathrm{LD_i}$ and $\mathrm{LD_j}$ respectively represent the areas that two different learners are interested in learning; $\mathrm{E_i}$ and $\mathrm{E_j}$ respectively represent the current level of knowledge that two different learners have in the field they are interested in learning [22]. The relationship between learners and learning domains belongs to a one-to-one correspondence, as calculated in Eq. (9).

$$\eta_{\mathrm{ik}}(\mathrm{t})=\sum_{\mathrm{LUA_j}\in\mathrm{Sim(LA_i)}}\mathrm{Dis}(\mathrm{LA_i},\mathrm{LA_j})\times\mathrm{Count}(\mathrm{LA_j},\mathrm{LUA_k}) \tag{9}$$

In Eq. (9), $\eta_{\mathrm{ik}}(\mathrm{t})$ represents similarity; $\mathrm{LUA_k}$ represents the overall learning area; $\mathrm{LUA_j}$ represents the current learning field; $\mathrm{Sim(LA_i)}$ represents the learning group. The probability that the learner has previously learned knowledge in the learning field is shown in Eq. (10) [23].

$$P_{ik}(t)=\begin{cases}\dfrac{[\eta_{ik}(t)]^{a}}{\sum\limits_{k}[\eta_{ik}(t)]^{a}}\\ 0\end{cases} \tag{10}$$

In Eq. (10), $P_{ik}(t)$ represents the probability that the learner has learned the selected learning area within time t, the probability of not learning is 0, and the probability of learning is $\dfrac{[\eta_{ik}(t)]^{a}}{\sum\limits_{k}[\eta_{ik}(t)]^{a}}$; α represents the weight of similarity; $k*$ represents the number of learners and learning domains. The similarity estimation between learning units is shown in Eq. (11) [24].

$$\Phi_{jk}=\mathrm{Dis}(\mathrm{UA_j},\mathrm{UA_k})\times\mathrm{Dis}(\mathrm{C_j},\mathrm{C_k}) \tag{11}$$

In Eq. (11), UA_j and UA_k respectively represent the number of similar contents present in the learning unit, while C_j and C_k respectively represent the degree of correlations between learning neighborhoods. The representation of pheromone concentration is the core of path planning, and the calculation method is shown in Eq. (12) [25].

$$\varepsilon=(\Delta\tau_{jk}(t)-60)/(90-60) \tag{12}$$

In Eq. (12), when $\Delta\tau_{jk}(t)$ represents the pheromone variation value, and its range is between [60, 90]. ε represents the adjustment factor. According to the change value of pheromones, learners are scored, and the current learning path is adjusted using adjustment factors. The process of judging the current level of knowledge mastery of learners is shown in Fig. 5.

When recommending learning paths based on the initial learning state of learners, due to their low or zero knowledge level and inability to rely on the calculation of the previous path or pheromone, it is necessary to choose based on the degree of matching with the course database in the model. The recommendation for learners in the initial state is shown in Eq. (13).

$$p_{oi}=\sum_{r\neq o}\lambda_{or}\cdot\theta_{ri} \tag{13}$$

In Eq. (13), p_{oi} represents the similarity between the learner in the starting learning state and the learning unit; λ_{or} is the similarity between learner S_o and learner S_r; θ_{ri} is the similarity between learner S_r and learning unit L_i. When recommending to learners who are in an adaptive state and are satisfied with the current course, it is only necessary to recommend units based on the adjustment factor, as shown in Eq. (14).

$$p_{ij}^{S_o}(t)=\begin{cases}\dfrac{\left[\eta_{ij}\right]^{\alpha^*}\times\left[\tau_{ij}(t)^*\right]^{\beta^*}}{\sum\left[\eta_{ij}\right]^{\alpha^*}\times\left[\tau_{ij}(t)^*\right]^{\beta^*}}, L_j|\ \phi_{ij},, \varepsilon\\ 0, L_j|\ \phi_{ij}>\varepsilon\end{cases} \tag{14}$$

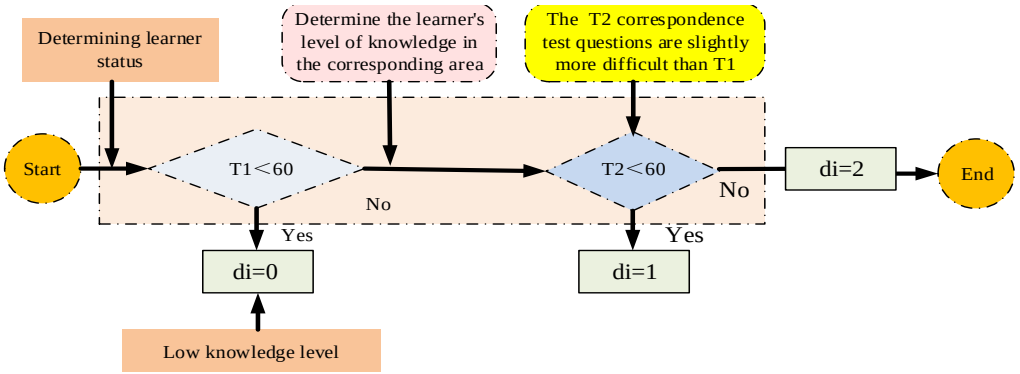


Fig. 5. Flowchart for judging the level of knowledgeable people.

In Eq. (14), $p_{ij}^{S_o}(t)$ represents the probability of S_o learners learning from L_i unit to L_j unit during t time; η_{ij} represents the pheromone between two units ij at time t; $\tau_{ij}(t)^*$ represents the pheromone density of the path connecting two units ij at t time; α^* and β^* respectively represent the influencing factors of path pheromone density and learners' level of knowledge mastery. ϕ_{ij} represents the degree of overlap between the feature values of two learning courses in ij. When recommending learners who have adjusted their status, it indicates that the learner is not satisfied with their current level of knowledge in the learning course and needs to revise their learning path, as shown in Eq. (15).

$$p_{ij}^{S_o}(t)=\begin{cases}\dfrac{\left[\eta_{ij}\right]^{\alpha^*}\times\left[\tau_{ij}(t)\right]^{\beta^*}}{\sum\left[\eta_{ij}\right]^{\alpha^*}\times\left[\tau_{ij}(t)\right]^{\beta^*}}, L_j \mid d_m^o = C_m^j \\ 0, L_j \mid d_m^o \neq C_m^j\end{cases} \quad (15)$$

In Eq. (15), d_m^o represents the degree of knowledge mastery of learner S_o in the knowledge domain n; C_m^j represents the breadth of knowledge of learning unit L_j within the knowledge domain n.

III. Results

To validate the performance of the proposed algorithm model, the dataset source and experimental platform were first introduced. Then relevant performance tests were conducted on the algorithms before and after optimization, as well as other algorithms, in different scenarios. Secondly, different functions were used to compare the proposed algorithm with other algorithms. Finally, the proposed algorithm was compared with other algorithms in terms of practical application effects.

A. Performance Testing of Intelligent Learning Model Based on ACO Algorithm

The experimental platform adopted the hardware configuration of Genuine Intel equipment ® CPU 2140 (dual core) 1.6GHz, 512MB memory. The experiment was conducted on the MATLAB 7.0 platform. The data used in the study was the TSPLIB dataset, and the related problems of this dataset achieved optimal solutions and were widely recognized. It selected relevant questions from this dataset to train and test the performance of the proposed algorithm. Combined with the comparison of three classic models in the past, the superiority of this method could be clarified. Classic algorithms included AS, MMAS, and ASrank. The study adopted the four core issues present in the TSPLIB dataset, as shown in Table I.

To verify the search capability of the proposed algorithm, four out of the seven shutdown key questions were selected for performance testing, and to compare it with three classic algorithms, the optimization results are shown in Fig. 6. From the figure, the proposed AS-N algorithm had better search ability and could search for the shortest path in a shorter time. The four algorithms ASrank, MMAS, AS, and NFACA converged to the optimal solution of problem Ulysses21 at 40, 32, 36, and 27, respectively, and the proposed results performed the best. The proposed NFACA algorithm reduced the optimal solution by 10% compared to the original AS algorithm, and reduced it by 12% and 24% compared to MMAS and ASrank. The performance of the four algorithms on datasets of four different problems is not significantly different, therefore the proposed algorithm has better reliability and convergence speed.

To verify the reliability of the proposed optimization, the actual application effects of the proposed NFACA algorithm were compared and validated with 10 dimensional test functions of ASrank, MMAS, and AS algorithms. The test functions were Sphere, Ellipse, Ackley, and Griebank, respectively. From Fig. 7, the proposed algorithm had the fastest convergence speed when tested in Sphere and Ellipse functions, with an average function value of 10-20, while the average function values of the other three algorithms were around 103. Under the influence of effective and critical subsets, its solution accuracy was higher, and when the adaptive function was estimated around 4000 times, it was already more pronounced on the average of the function than other algorithms. In the Ackley case, all four algorithms achieved good function averages, but NFACA still reached 104 before one-third of the optimization process. During the Griewankf test, only the optimized NFACA algorithm avoided the problems of "premature convergence" and getting stuck in local optima. When the number of adaptive evaluations was between 0-3000, the convergence speed was significantly faster than other algorithms.

TABLE I. Four Problems with the TSPLIB Dataset

Question title	Problem description	Optimal solution
Ulysses21	Odyssey of ulysses	78
Att44	44 cities of the US	33117
Berlin50	50 location in Berlin	7033
Eril21	21 problem of city	519
KroA100	100 cities in Kroatien	21,282
Lin105	105 cities with global optimal	14,379
Ch130	130 cities in Switzerland	6434

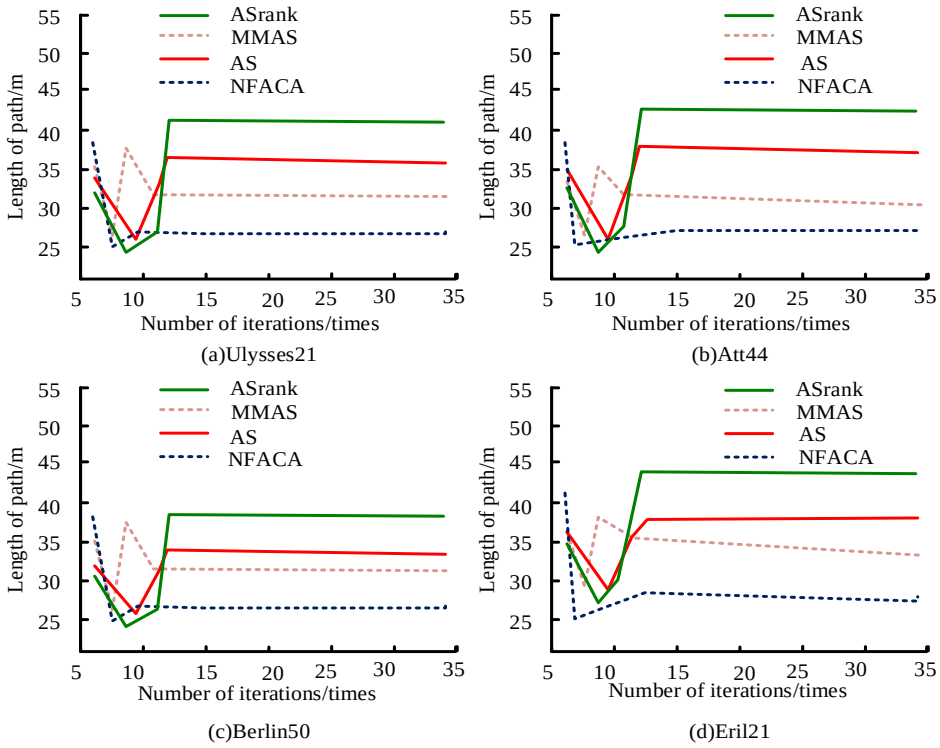


Fig. 6. Comparison of search ability of four algorithms in different problems.

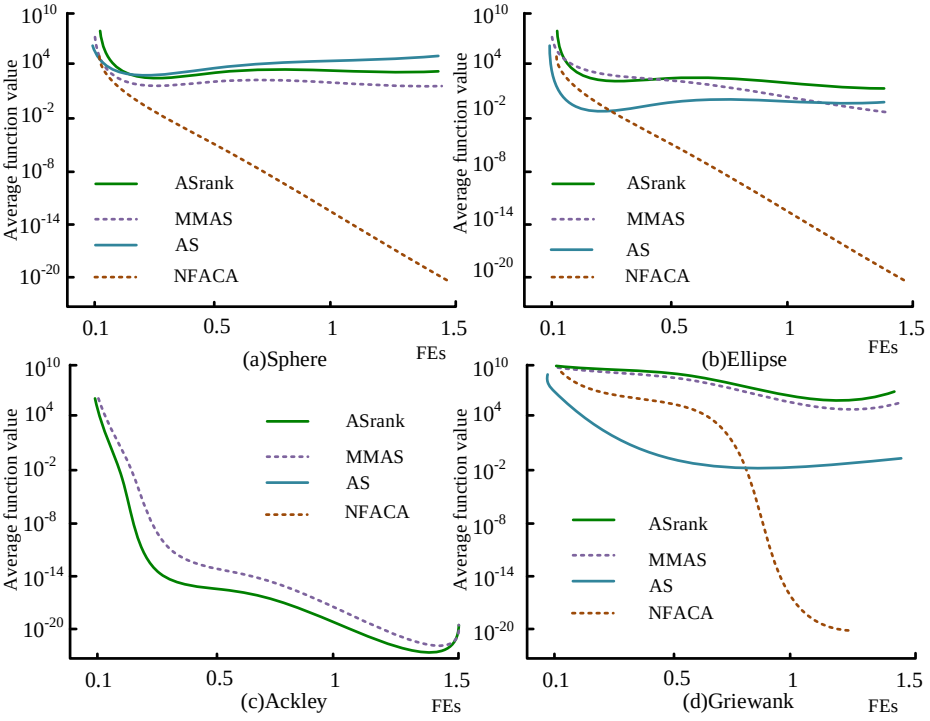


Fig. 7. Comparison of convergence curves of each algorithm in four functions.

B. *Analysis of the Actual Effect of the Intelligent Learning Model Based on the ACO Algorithm*

The convergence process curve of the proposed model with increasing iteration times was analyzed, the application effect of the model in four different real environments was evaluated, and it was compared with MMAS and AS models. From Fig. 8, all four models began to converge after more than eight searches. The ACO algorithm had a better convergence effect, with the shortest path of 53.5. The AS model had the worst convergence effect, with a difference of 5% compared to the ACO algorithm. The convergence degree of

the MMAS model and the proposed ACO algorithm was not significantly different, but they began to converge at the 6th search, showing a phenomenon of "premature convergence". The proposed algorithm achieved the best convergence effect in all four environments, indicating that it can stably and effectively search for the optimal solution in different complex environments.

To verify the stability of the constructed intelligent learning model in the application of real learning course recommendation scenarios, the stability of the proposed method, ACO and MMS algorithms were compared under different RMSE and RD. The results are shown in Fig. 9. The stability of all three methods decreased with the increase of RMSE, but when RMSE reached 6%, the stability of the proposed method no longer decreased but tended to stabilize. Among them, the ACO algorithm showed the greatest decrease, with a total decrease of 12%. The stability decline of MMS algorithm slowed down in the later stage, with a total decrease of 10%. As the absolute value of the maximum RD increased, the stability of ACO and MMS algorithms has been on a downward trend, decreasing by 12% and 15% respectively. However, the stability of the proposed algorithm was the best, and when RD increased to 8%, the stability did not continue to decrease but increases, with a total decrease of only 5%. The results indicate that the proposed method has higher stability and is more reliable for clustering classification in practical employment recommendation scenarios.

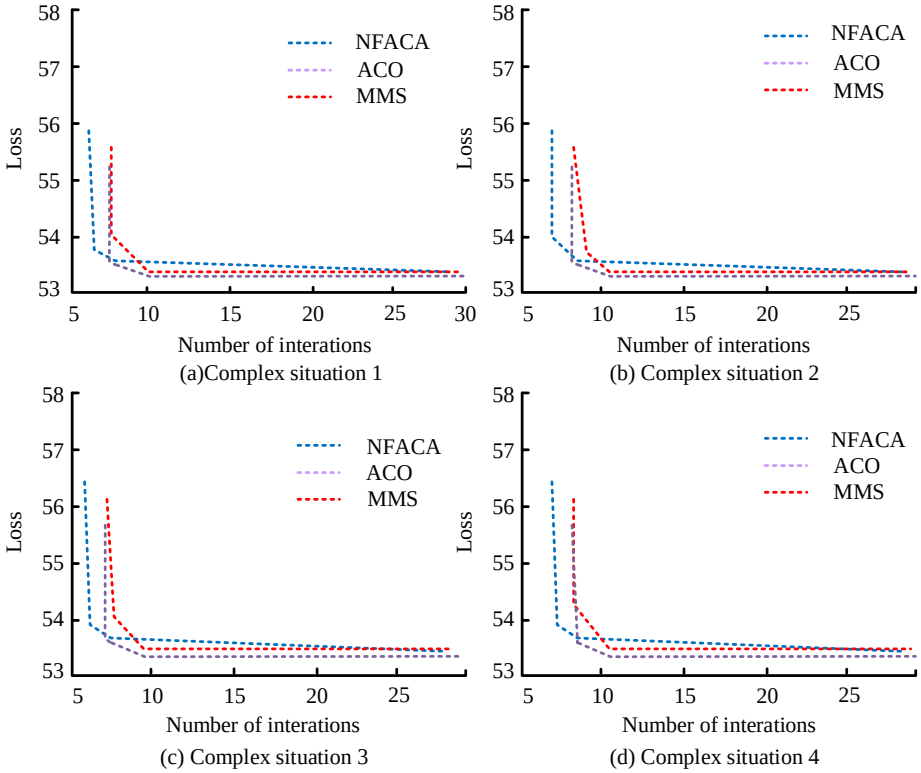


Fig. 8. Comparison of different algorithms for path search.

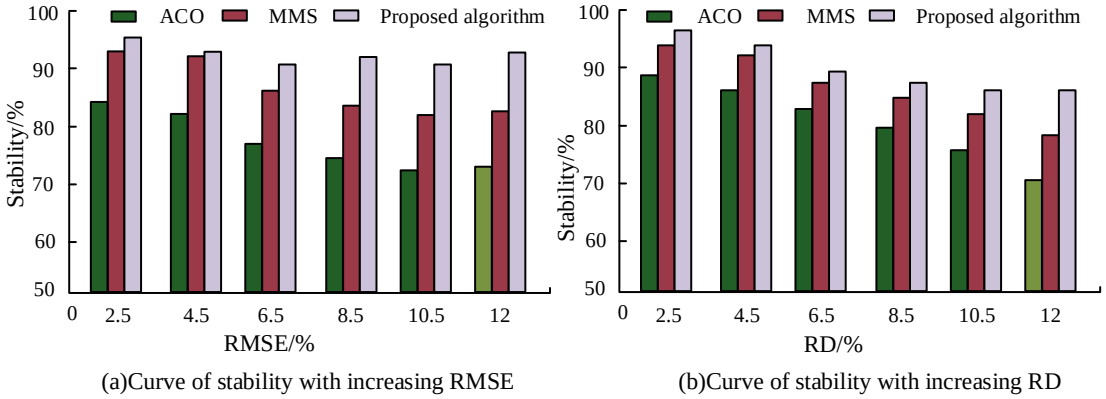


Fig. 9. Changes in stability by method.

To test the accuracy of calculating similarity between students using NFACA model, it was applied to three different university course learning platforms in real scenarios for simulation experiments, and the effects were compared using AS, TaSimRank, and RD3 algorithms. The results are shown in Fig. 10. From Fig. 10, in the datasets of the three scenarios, the accuracy of all four models increased with the increase of k. Among them, the accuracy of the proposed model achieved the best results in three scenarios, with differences within 0.2 in different datasets, and was on average 10%, 11%, and 13% higher than the accuracy of AS, TaSimRank, and RD algorithms, respectively. The RD algorithm had the lowest variation in calculation accuracy among different university datasets, ranging from 0.18 to 0.7. So, the proposed model has higher accuracy and can achieve good repeatability when applied in different course recommendation scenarios.

To test the similarity calculation efficiency of the intelligent learning model proposed by the research in recommending learning units to learners, the ACO and MMS algorithms were compared in three different university learning course recommendation scenarios under two different paths, Naive and Block. The results are shown in Fig. 11. From the figure, the proposed model used the least amount of computation time in both the path Block and path Naive scenarios. Among them, the Naive scenario took the least amount of time, with an average of only 1025s in three different scenarios, which was 16% and 23% lower than the ACO and MMS algorithms, respectively. In the path Block scenario, the proposed model still took the least amount of time, averaging only 1302 seconds, which was 12% and 19% lower than the ACO and MMS algorithms, respectively. From the results, the proposed algorithm performs the best in terms of computation time under different path comparison times. Therefore, this method is applicable in practical environments.

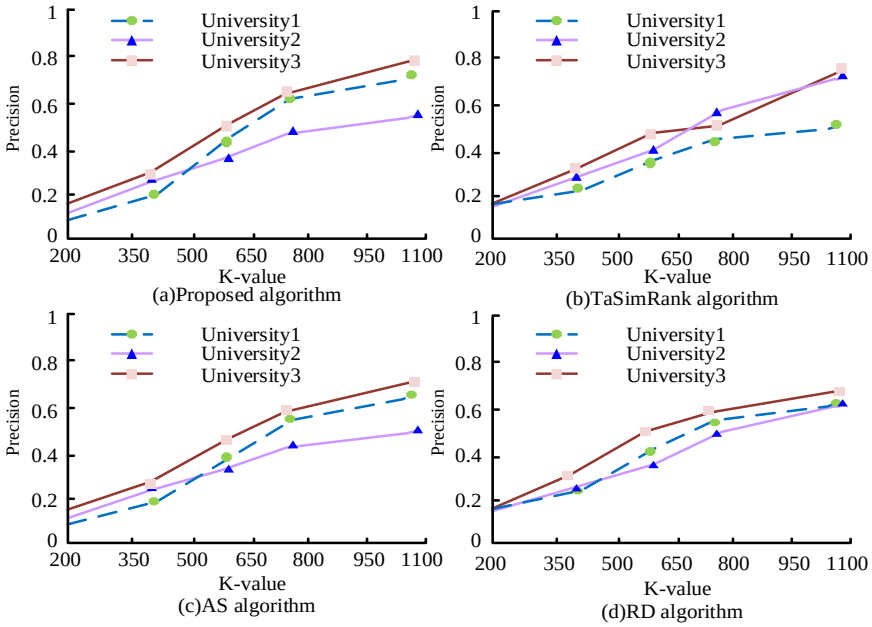


Fig. 10. Comparison of computational accuracy of different algorithms in different scenarios.

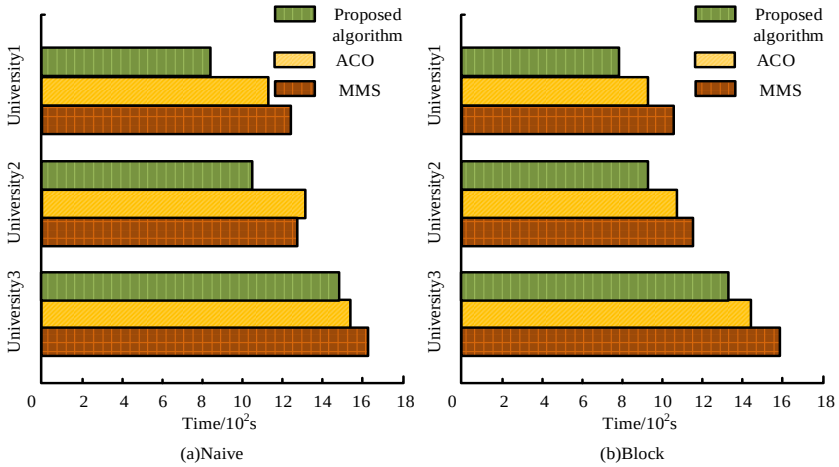


Fig. 11. Comparison of computational accuracy of different algorithms in different scenarios.

IV. Discussion and Conclusion

The rapid development of online learning has brought many benefits, but it is also accompanied by problems that need to be solved. In the field of course recommendation, although existing intelligent models have shown some effectiveness, they still face limitations in terms of real-time and accuracy. Liu Y and other researchers proposed a pseudo-random algorithm for path planning, which solves the problem of local optima. However, when facing large-scale datasets, there are limitations to their computational efficiency and scalability. Researchers such as Zhang G have developed an ant colony algorithm based on a genetic mutation algorithm to improve the accuracy and applicability of meteorological route planning. However, when directly applied to the field of course recommendation, the complex parameter adjustment and domain specificity issues limit its universality. Mousavinasab et al. proposed a machine learning model that better classifies learners, although it has shown good personalized recommendation performance in physics, chemistry, and clinical fields. However, how to accurately capture learners' interest changes and dynamic needs in course recommendations is still an unsolved problem. Given the limitations of the aforementioned research, an innovative intelligent learning model based on a negative feedback ant colony algorithm is proposed. Compared to traditional methods, the model proposed by the research institute has shown significant advantages in reducing the time cost for learners to choose courses, improving recommendation accuracy, and enhancing user experience.

The results showed that the proposed model achieved a convergence speed accuracy of 10^{-20} when tested in Sphere and Ellipse functions, and the optimal solution for problem Ulysses21 was 27, all of which were better than ASrank, MMAS, and AS. The NFACA mechanism had better convergence effect than ASrank, MMAS, and AS algorithms, with a shortest path of 53.5. After reaching RMSE and RD distributions of 6% and 8%, the stability of the proposed method no longer decreased with increasing RMSE. In the scenarios of path Block and path Naive, the proposed algorithm had an average computation time of only 1011, which was better than the ACO and MMS models. In the path Block scenario, the computation time of the three algorithms was slightly reduced compared to the Naive path scenario. This was because the Naive path had more comparisons, which greatly slowed down the algorithm's running speed and increased computation time. In datasets from three different university scenarios, the accuracy of all four models increased with the increase of k. Among them, the accuracy of the proposed model achieved the best results in three scenarios, and the difference in accuracy between different datasets was within 0.2, which was 10%, 11%, and 13% higher than the average accuracy of AS, TaSimRank, and RD algorithms, respectively. Therefore, the proposed method provides a new solution for path planning in the field of online course recommendation, increasing convergence speed while conducting global search, and improving the accuracy, stability, and applicability of course recommendation platform algorithms. However, there are certain limitations in the methodology, as there are multiple parameters in the negative feedback ant colony algorithm, including pheromone heuristic factor, expected heuristic factor, and pheromone evaporation coefficient. Adapting these parameters to further enhance the algorithm's ability to adaptively adjust the environment can be a future research direction. It can also be combined with a differential particle optimization algorithm to improve the diversity of solutions.

References

[1] Labrague LJ, De los Santos JA. Transition shock and newly graduated nurses' job outcomes and select patient outcomes: A cross-sectional study. Journal of nursing management, 2020 ,28(5):1070-1079.

[2] Li J, Netessine S. Higher market thickness reduces matching rate in online platforms: Evidence from a quasiexperiment. Management Science, 2020 ,66(1):271-289.

[3] Simone II CB, Bogart JA, Cabrera AR, Daly ME, DeNunzio NJ, Detterbeck F, Faivre-Finn C, Gatschet N, Gore E, Jabbour SK, Kruser TJ. Radiation therapy for small cell lung cancer: an ASTRO clinical practice guideline. Practical radiation oncology, 2020 ,10(3):158-173.

[4] Hebbi C, Mamatha H. Comprehensive Dataset Building and Recognition of Isolated Handwritten Kannada Characters Using Machine Learning Models. Artificial Intelligence and Applications, 2023, 1(3):179-190.

[5] Bandewad G, Datta K P, Gawali B W, & Pawar, S. N.Review on Discrimination of Hazardous Gases by Smart Sensing Technology.Artificial Intelligence and Applications,2023, 1(2): 86-97.

[6] Gao W. New ant colony optimization algorithm for the traveling salesman problem. International Journal of Computational Intelligence Systems, 2020 ,13(1):44-55.

[7] Liu Y, Cao B, Li H. Improving ant colony optimization algorithm with epsilon greedy and Levy flight. Complex & Intelligent Systems,2021 ,7(4):1711-1722.

[8] Stodola P. Hybrid ant colony optimization algorithm applied to the multi-depot vehicle routing problem. Natural computing, 2020 ,19(2):463-475.

[9] Zhang G, Wang H, Zhao W, Guan Z, Li P. Application of improved multi-objective ant colony optimization algorithm in ship weather routing. Journal of Ocean University of China, 2021(6) ,20:45-55.

[10] Li W, Wang GG, Gandomi AH. A survey of learning-based intelligent optimization algorithms. Archives of Computational Methods in Engineering, 2021 ,28(5):3781-3799.

[11] Patibandla RS, Veeranjaneyulu N. A SimRank based ensemble method for resolving challenges of partition clustering methods. Journal of Scientific & Industrial Research,2022 ,79(4):323-327.

[12] Aggarwal K, Mijwil MM, Al-Mistarehi AH, Alomari S, Gök M, Alaabdin AM, Abdulrhman SH. Has the future started? The current growth of artificial intelligence, machine learning, and deep learning. Iraqi Journal for Computer Science and Mathematics,2022 ,3(1):115-123.

[13] Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. Electronic Markets,2021 ,31(3):685-695.

[14] Yang K, Shi Y, Zhou Y, Yang Z, Fu L, Chen W. Federated machine learning for intelligent IoT via reconfigurable intelligent surface. IEEE network, 2020 ,34(5):16-22.

[15] Mousavinasab E, Zarifsanaiey N, R. Niakan Kalhori S, Rakhshan M, Keikha L, Ghazi Saeedi M. Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. Interactive Learning Environments,2021 ,29(1):142-163.

[16] Fang X, Heldring A, Rius J M, Chen X, Cao Q. Nested fast adaptive cross approximation algorithm for solving electromagnetic scattering problems. IEEE Transactions on Microwave Theory and Techniques, 2020, 68(12): 4995-5003.

[17] Deng W, Xu J, Song Y, et al. An effective improved co-evolution ant colony optimisation algorithm with multi-strategies and its application. International Journal of Bio-Inspired Computation, 2020, 16(3): 158-170.

[18] Sadiq A T, Raheem F A, Abbas N. Ant colony algorithm improvement for robot arm path planning optimization based on D strategy.

International Journal of Mechanical & Mechatronics Engineering, 2021, 21(1): 196-111.

[19] Kanso B, Kansou A, Yassine A. Open capacitated ARC routing problem by hybridized ant colony algorithm. RAIRO-Operations Research, 2021, 55(2): 639-652.

[20] Wang Z, Ding H, Li B, Bao L, Yang Z, Liu Q. Energy efficient cluster based routing protocol for WSN using firefly algorithm and ant colony optimization. Wireless Personal Communications, 2022, 125(3): 2167-2200.

[21] Lv Z, Chen D, Feng H,Wei W, Lv H. Artificial intelligence in underwater digital twins sensor networks. ACM Transactions on Sensor Networks (TOSN), 2022, 18(3): 1-27.

[22] Liu Y, Cao B. A novel ant colony optimization algorithm with Levy flight. Ieee Access, 2020 ,8:67205-67213.

[23] Too J, Abdullah A R. A new and fast rival genetic algorithm for feature selection. The Journal of Supercomputing, 2021, 77(3): 2844-2874.

[24] Wang B, Liu Z, Li Q, Prorok A. Mobile robot path planning in dynamic environments through globally guided reinforcement learning. IEEE Robotics and Automation Letters, 2020, 5(4): 6932-6939.

[25] Yang Y, Juntao L, Lingling P. Multi-robot path planning based on a deep reinforcement learning DQN algorithm. CAAI Transactions on Intelligence Technology, 2020, 5(3): 177-183.

Automated Detection of Malevolent Domains in Cyberspace Using Natural Language Processing and Machine Learning

Saleem Raja Abdul Samad[1], Pradeepa Ganesan[2], Amna Salim Al-Kaabi[3], Justin Rajasekaran[4], Singaravelan M[5], Peerbasha Shebbeer Basha[6]

IT Department, College of Computing and Information Sciences, Shinas, University of Technology and Applied Sciences, Sultanate of Oman[1, 2, 3, 4]

Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R & D Institute of Science and Technology, Chennai, Tamil Nadu, India[5]

Department of Computer Science, Jamal Mohammed College, Tiruchirappalli, Tamil Nadu, India[6]

***Abstract*—Cyberattacks are intentional attacks on computer systems, networks, and devices. Malware, phishing, drive-by downloads, and injection are popular cyberattacks that can harm individuals, businesses, and organizations. Most of these attacks trick internet users by using malicious links or webpages. Malicious webpages can be used to distribute malware, steal personal information, conduct phishing attacks, or perform other malicious activities. Detecting such malicious websites is a tedious task for internet users. Therefore, locating such a website in cyberspace requires an automated detection tool. Currently, machine learning techniques are being used to detect such malicious websites. The majority of recent studies derive a limited number of features from webpages (both benign and malicious) and use machine learning (ML) algorithms to detect fraudulent webpages. However, these constrained capabilities might not use the full potential of the dataset. This study addresses this issue by identifying malicious websites using both the URL and webpage content features. To maximize detection accuracy, both ngrams and vectorization methods in natural language processing are adopted with minimum feature-set. To exploit the full potential of the dataset, the proposed approach derives the 22 common linguistic features of the URL and generates ngrams from the domain name of the URL. The textual content of the webpages was also used. The research employs seven machine learning algorithms with three vectorization methods. The outcome reveals that the proposed method outperformed the results of previous studies.**

***Keywords*—Machine learning; N-gram; linguistic features; natural language processing (NLP); malicious webpage**

I. Introduction

A malicious webpage is a website intentionally crafted to harm users or their devices, often orchestrated by hackers and cybercriminals utilizing various techniques to exploit vulnerabilities. These webpages pose significant threats, serving as conduits for malware distribution, personal information theft, phishing attacks, and other malicious activities. Cybercriminals employ common techniques like phishing, where deceptive URLs (Uniform Resource Locators) or webpages mimic legitimate sites to trick users into divulging login credentials or personal data, facilitating identity theft or financial fraud. Another tactic involves drive-by downloads, concealing malicious code within a webpage to exploit vulnerabilities in users' browsers or operating systems, and installing malware without their knowledge. Malvertising utilizes legitimate online advertising networks to redirect users to malicious pages, aiming to infect their computers with malware. Cryptojacking involves hijacking a user's computer processing power to mine cryptocurrency for attackers. Once users access a malicious webpage, they risk downloading malware, suffering personal information theft, or falling victim to various attacks. The consequences extend to severe impacts on individuals, businesses, and organizations, encompassing data theft, financial losses, reputational damage, and even physical harm.

The cost of cyber-attacks continues to rise and is a significant concern for organizations across all industries. According to IBM's 2024 Cost of a Data Breach Report, the global average cost of a breach has increased by 10% since 2023, reaching $4.88 million [1]. Global cybercrime expenditures are anticipated to increase from $3 trillion in 2015 to $10.5 trillion annually by 2025, according to Cybersecurity Ventures research [2]. Phishing attacks, which targeted 1339 brands, accounted for 36% of US data breaches in 2023. It became the second-most expensive source of compromised credentials with 5 million unique phishing sites, signaling a significant change in the dynamics of breaches across industries [3]. The financial impact of a cyber-attack depends on factors such as its severity, organizational size, and the compromised data. To mitigate these impacts, proactive measures like implementing robust security systems and conducting employee training programs are essential. While conventional security systems often rely on stored lists of URLs to block harmful links, recent attacks exploit vulnerabilities using short URLs and algorithmically generated ones, evading traditional security measures. Consequently, the necessity for an automated system employing machine learning techniques becomes increasingly critical to detect and prevent these evolving cyber threats effectively. Timely detection plays a pivotal role in ensuring a proactive defense against cyber

Funding: This research work is supported by the Research Internal Funding Program (RIFP) 2024, of the University of Technology and Applied Sciences-Shinas, Sultanate of Oman.

threats, effectively safeguarding users and their devices from various malicious activities.

Machine learning-based detection relies on feature engineering. The researcher identifies and uses six types of features for the detection of malicious webpages such as:

- Linguistic analysis of URL [3][4]: URL consists of several parts such as protocol, domain name, path, query-string, etc. Using these parts for the detection process such as a number of subdomains in the URLs, protocol used in the URL length of the path, etc.
- Information obtained Domain Name System (DNS) [3][5][6]: DNS is a critical component of the internet infrastructure, and it plays an important role in the functioning of URLs. To generate features, researchers obtain information from the DNS server such as server location, registration and expiry date of the website, etc.
- Structural information of webpage [7]: The structural information of a webpage refers to the underlying organization of the webpage's content, including its layout, formatting, and elements of the page. For malicious website detection, the researcher analyzes HTML tags, DOM objects, CSS, and JavaScript elements, and other web page components.
- Linguistic analysis of webpage contents [8]: This is the written content of the webpage, including headings, paragraphs, and other text elements. The linguistic analysis involves character and word-level analysis to find harmful websites, such as finding probable keywords, tokens, and keyword weight, etc.
- Website Reputation [9]: Website reputation or ranking refers to the perceived level of trustworthiness, authority, and relevance of a particular website, as assessed by search engines, users, and other entities. Websites with a high reputation or ranking are generally viewed as more credible and valuable and may receive more traffic and engagement as a result. Ranking of the website globally and locally is considered as a feature for malicious webpage detection.
- Visual Similarity of the Webpage [10]: Visual similarity of a webpage can be a useful tool in malicious webpage detection. Malicious actors may attempt to create webpages that are visually similar to legitimate sites to trick users into providing sensitive information or installing malware. One approach to detecting visual similarity in webpages is to use computer vision techniques such as image processing, feature extraction, and machine learning. These techniques can be used to identify and compare visual elements of webpages, such as logos, fonts, and layouts.

Among all these features, researchers preferably use the linguistic analysis of URLs because it doesn't involve any risks. Web page content, on the other hand, gets less attention due to its risky nature. However, in comparison, webpage contents are a rich source of malicious webpage detection. The majority of recent studies derive a limited number of features from the URL and use machine learning algorithms (ML) to detect fraudulent webpages. However, these constrained capabilities might not use the URL's full potential. This study addresses this issue by identifying malicious websites using both the URL and webpage content features. To maximize detection accuracy with a minimum hybrid feature set, the proposed approach uses both URL and webpage contents. To exploit the full potential of the dataset, the proposed approach derives the 22 common linguistic features of the URL and generates ngrams from the domain name of the URL. Moreover, it extracts text from the webpage's paragraph tags. In the experiment, three distinct vectorization techniques (Count vectorizer, TFIDF, and Hashing vectorizer) are used to convert text into real numbered 2-D vectors. Lastly, the vectorized features and linguistic features are integrated to generate the experimental dataset. The incorporation of vectorized contents with URL linguistic features (hybrid approach) eliminates the necessity for generating and selecting a limited number of features. This ensures the full exploitation of the dataset's potential. The research employs seven machine learning algorithms. The outcome reveals that the proposed method outperformed the results of previous studies.

Contribution of the paper:

- Both URL and Webpage contents are utilized to generate features.
- Two different feature sets are generated using URLs of the webpages namely linguistic features and character-level ngrams of the URL
- Textual contents are extracted from the paragraph tags of the webpages.
- Three different NLP methods (Count vectorizer, TFIDF, and Hashing vectorizer) are used to vectorize the textual contents such as ngrams of the URL and textual contents from the webpages.
- NLP methods are used for feature generation instead of manual feature extraction.
- A balanced dataset (malicious and benign) is used for the experiments.
- Seven machine learning algorithms are evaluated for better detection accuracy.
- Achieve higher accuracy and f1-score with minimum feature set.

The structure of this paper is organized as follows. Section I outlines the significance of the problem and an overview of the solution strategy. Existing research is presented in Section II. The NLP techniques and machine learning classifiers are presented in Section III. Section IV presents the proposed method, and Section V reports the experimental results. Section VI provides a conclusion of the paper.

II. Related Works

Malicious webpage detection is a critical component of web security, as it can help to protect users from a variety of online threats, such as malware, phishing, and other types of

malicious activity. Extensive research has been conducted in this area. Table I summarizes recent research studies in this field. Saleem et al. [7] introduced a technique for classifying malicious webpages by analyzing their content through machine learning and deep learning approaches. The presented method exclusively uses of tags, events, keywords, scripts in the website for classification. The study utilizes the Kaggle dataset. From the dataset, over 206 features are retrieved. The selectKbest approach selects the highest-scoring features for the experiment. Scores in selectKbest are based on chi-square tests. Support vector machine (SVM), random forest (RF), and convolutional neural network (CNN) are used to test the system. The results demonstrate that random forest and support vector machine obtains the accuracy of 93% and 88% respectively. Saleem et al. [11] suggested a simple technique for detecting malicious URLs. The authors note that traditional approaches to malicious URL detection, which rely on blacklists and reputation-based systems, are not effective against new or previously unseen threats. To address this, the authors propose a new approach that uses a set of lexical features, including the length of the URL, the presence of certain characters, and the number of subdomains, to train a machine learning model to detect malicious URLs. The authors evaluate their method using a UNB dataset containing over 68851 URLs, including both malicious (33473) and benign (35378) URLs. From the dataset, 27 lexical features were taken, and 20 of those features were used in the experiment. The study's findings indicate that the suggested method achieves high accuracy in detecting malicious URLs, with 99% accuracy for random forest (RF) and 98% accuracy for k-nearest neighbor (k-NN) algorithms, respectively. Malak et al. [12] evaluate different feature sets for detecting malicious URLs using machine learning and deep learning models. The authors note that existing approaches to malicious URL detection often rely on a single feature set, which may limit their effectiveness. To address this, the authors evaluate three different feature sets: lexical, network-based, and content-based. Many feature selection procedures, including correlation analysis, ANOVA, and chi-square, were utilized to select features from the dataset. Naive Bayes (NB) was shown to be the most appropriate method for identifying malicious URLs in the used dataset, with an accuracy of 96%. Kamel [13] proposes a new approach for detecting and analyzing phishing attacks on social networks, specifically on Twitter. The authors note that phishing attacks on social networks are a growing problem, and traditional approaches to detecting them, such as manual analysis or keyword-based filtering, are not effective. To address this, the authors proposed a new approach by using machine learning algorithms to identify phishing attacks on Twitter. The approach involves analyzing the content of tweets and using machine learning to classify them as either phishing or non-phishing. For this experiment, a UCI phishing, Phishtank, and MillerSmiles datasets were utilized, and roughly 25 features were retrieved from the URL, webpage, and DNS server. For the studies, LR, SVM and RF were applied, yielding 90.28%, 93.43%, and 95.51% accuracy, respectively. Lakshmanarao et. al [14] proposes a web application for detecting malicious URLs using natural language processing (NLP) and machine learning techniques. This method makes use of a count, a TFIDF, and a hashing vectorizer. The outcome demonstrates that hashing vectorizer and random forest obtained 97.5% accuracy. Machine and deep learning help detect email-delivered malicious URLs, according to Joshi et al [15]. To classify URLs, the proposed method relied on their lexical features. The experiment employed a dataset from openphish, alexa, and fire eye. The proposed method used 23 distinct lexical characteristics to distinguish between malicious and benign URLs. Combining the lexical features with 1000 trigram-based features yielded 1023-long numerical vectors to represent the URLs. To identify the most important traits, correlation and scatter matrices were used. The results show that 92% accuracy is achieved by the random forest method. Hong et al. [16] employ lexical analysis and feature quantification to identify potentially dangerous domain names. The method comprises of two stages for accurate and successful detection. A domain name is compared in the first step to a blacklist of harmful domain links. The degree to which the domain name modifications closely resemble the blacklist determines whether it is malicious or possibly malicious. In the second step, a suspected malicious domain name is assessed using the reputation value of an N-gram model. Using the N-gram approach, a collection of whitelist/blacklist substrings is extracted from the top 100,000 regular Alexa domain names. The frequency of substrings on whitelists and blacklists determines their weights. Lastly, the authenticity of the possibly dangerous domain name is determined by its reputation value. The outcome shows that the accuracy rate for the proposed method, LA-FQ, is 94.16%. Josh et al. [17] used random forest to detect algorithmically generated domains. For testing, the dataset comprised regular and algorithmically generated domain names from several malware families. Masked N-grams and other domain name data were extracted. Results show that masked N-grams provide improved performance and detection accuracy compared to state-of-the-art methods. To detect harmful web links using NLP, Saleem et al. [18] suggested an ensemble classifier. The author observed that feature generation needs effort and topic expertise. Sometimes the generated features don't maximize the data set. Hence, the suggested method generates a feature set from the URL using the NLP method and classifies using an ensemble classifier. Two datasets (D1 and D2) were used in the experiment. D1 and D2 had 91.4% and 99.1% accuracy, respectively. The phishing URL detection method used by Ozgur et al. [19] is based on machine learning and a natural language processing (NLP) feature set. URLs were broken up and features were taken from them. For the experiment, seven different machine learning algorithms were used. The author states that the proposed method is language-independent and uses a large dataset. The outcome demonstrates that random forest with NLP features has an accuracy of 97.98%.

TABLE I. SUMMARY OF THE EXISTING WORK

Author	Dataset/Data Source	Features	Accuracy	Issues
Saleem et al. [7]	Kaggle dataset	Web content-based features	SVM: 88%, RF: 93%	URL-based features and webpage textual contents were ignored.
Saleem et al. [11]	UNB Dataset	Lexical features of URL	RF:99%, KNN:98%	Limited number of features. Web content features were ignored
Malak et.al [12]	Singh & Kumar Dataset	Lexical, network-based, and content-based features	NB:96%	Limited features. The full potential of the dataset was not used.
Kamel et.al [13]	UCI phishing, Phishtank, MillerSmiles datasets	Lexical, content based and DNS server features	LR: 90.28%, SVM: 93.43%, RF: 95.51%	Limited features. The full potential of the dataset was not used.
Lakshmanarao et.al[14]	Kaggle dataset	NLP -Vectorizer Methods	Hashing vectorizer with RF: 97.5%	Potentials of web contents were not utilized.
Joshi et.al [15]	Openphish, Alexa, Fire eye	Lexical Features & N-Gram Method	RF:92%	Web content features were ignored.
Hong et al. [16]	Alexa, Anquan organization, Malwaredomains.com, Malicious domain list, Zeus Tracker, Conficker, Torping, Symmni, PhishTank	N-gram Method	LA-FQ:94.16%	Web content & Lexical URL features were ignored.
Josh et.al [17]	Alexa, Bader repo extended	N-Gram Method	RF: 98.91%	Web content & Lexical URL features were ignored.
Saleem et.al[18]	URL, UNB, Phistank dataset	NLP -Vectorizer Methods	Weighted Soft Voting Classifier: D1:91.4% , D2:99.1%	Web content & Lexical URL features were ignored.
Ozgur et.al [19]	Phistank dataset, Yandex Search API	NLP based features	RF: 97.98%	Web content and other features were ignored.

III. BACKGROUND

The URL of the webpage and the webpage's content are both valuable resources for webpage classification, particularly for the detection of malicious webpages. A URL is a string that is used as a specific identifier to find resources online. It is made up of several components as shown in Fig. 1. Due to its risk-free nature, researchers employ the URL of the webpage for detection. Two distinct methods are used to extract features from the URLs.

1) Numerical/presence measurement: check the URL for the presence of specific words or characters and count the occurrence of the specific word or character, URL/domain/path length, etc.

2) Decompose the URL string using the ngram method and identify interesting features such as the amount of 2-grams and the average length of 2-gram tokens, etc.

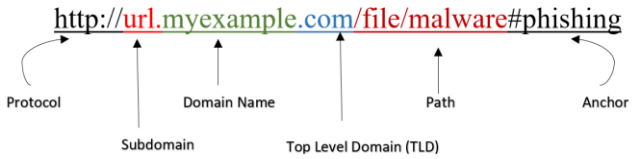


Fig. 1. Components of an URL.

Malicious websites can include a wide range of content that is intended to trick or exploit users, including fake login pages that steal login credentials, malware links that download malware onto the victim's computer, phishing forms that steal personal information, pop-up windows that direct users to phishing sites, redirect links that open phishing sites, malware scripts, fake surveys, and more. The textual content of a website serves as source for malicious activities. Generally, machine learning models operated on numerical features that are extracted from the raw dataset (list of url) and represented as 2-D array. However, in addition to a collection of URL features, this paper vectorizes URL text using natural language processing (NLP) methods. NLP has a set of vectorization methods. such as Count, Hashing, and TFIDF vectorizer. Vectorization converts text into numerical features. Then the vectorized data will be used in machine learning algorithms. Table II lists the machine learning algorithms used for the experiment.

a) Count vectorizer [20]: It works by first tokenizing the text corpus into individual words or n-grams (contiguous sequences of words). It then creates a vocabulary of unique tokens in the corpus and assigns an index to each token in the vocabulary. Finally, it counts the occurrence of each token in each document in the corpus and constructs a matrix of token counts, with one row per document and one column per token. The resulting matrix can be used as input to a machine learning algorithm for various NLP tasks.

b) TFIDF [20]: It measure the importance of each term in a document or a corpus. It considers both the frequency of a term in a document and the frequency of the term across the entire corpus. The basic idea behind TF-IDF is that a term is considered important if it appears frequently in a document but is not so common that it appears in every document in the corpus. The TF-IDF score of a term in a document is calculated by multiplying the term frequency (TF) of the term in the document by the inverse document frequency (IDF) of the term in the corpus.

The TF-IDF formula (1,2) can be expressed as follows:

$$TF(t,d) = (Number\ of\ times\ term\ t\ appears\ in\ document\ d) / (Total\ number\ of\ terms\ in\ document\ d) \quad (1)$$

$$IDF(t,D) = log\_e(Total\ number\ of\ documents\ in\ the\ corpus / Number\ of\ documents\ with\ term\ t\ in\ it)$$

$$TF-IDF(t,d,D) = TF(t,d) * IDF(t,D) \quad (2)$$

$where$:

- t: $a\ term$
- d: $a\ document$
- D: $the\ corpus$

c) Hashing vectorizer [20]: It is a technique used to convert a collection of text documents to a matrix of token occurrences. It is similar to Count-Vectorizer in that it creates a document-term matrix, but uses a hash function to convert each token to a numerical index, rather than storing the tokens as strings. This can be useful in situations where the vocabulary is very large, and the memory requirements of storing all the tokens are prohibitively high. The Hashing-Vectorizer works by mapping each token to a numerical index using a hash function. The resulting numerical index is then used as the column index in the document-term matrix. The size of the matrix is fixed in advance and depends on the number of features (i.e. the number of hashed tokens) that are required. This means that the hashing-vectorizer is a stateless transformer, and does not need to keep track of the mapping between tokens and numerical indices.

TABLE II. MACHINE LEARNING ALGORITHMS [21]

Algorithm	Explanation
Logistic Regression (LogR)	The logistic function, also referred to as the sigmoid function, is applied to a linear combination of the input variables via the logistic regression procedure. Any input value is converted by the logistic function to a probability value between 0 and 1. The method predicts the positive class (i.e., 1) if the probability is greater than a predetermined threshold (often 0.5), otherwise it predicts the negative class (i.e., 0).
Gaussian Naive Bayes (GNB)	Gaussian Naive Bayes is a variant of the Naive Bayes algorithm in machine learning that assumes that the features follow a Gaussian (normal) distribution. In Gaussian Naive Bayes, the likelihood of the features given the class is modeled as a normal distribution with mean μ and standard deviation σ for each feature and class.
Decision Tree (DT)	The Decision Tree algorithm iteratively splits data by feature or attribute value, creating a tree-like structure. At each node in the tree, the algorithm chooses the feature that delivers the best split based on some criterion, such as the Gini index or entropy. The procedure is repeated until a stopping requirement, such as a maximum depth or minimum number of samples per leaf, is fulfilled.
K Nearest Neighbors (KNN)	The number of nearest neighbors to take into account for each prediction is the initial step in the KNN algorithm's operation. The algorithm then determines the k training samples that are most similar to each new input based on a distance metric (e.g., Euclidean distance or Manhattan distance). The algorithm then calculates a forecast by averaging the target values of the k-nearest neighbors (for regression) or obtaining the majority vote (for classification).
Random Forest (RF)	Random Forest is a popular machine learning algorithm for both classification and regression tasks. It belongs to the family of ensemble learning methods, which combine multiple individual models to improve the overall predictive performance.
Gradient Boosting (GB)	Gradient Boosting is an ensemble learning method that combines weak models, usually decision trees, to generate a more accurate model. Each decision tree in the method is trained to rectify the faults of the previous tree. All tree predictions are added to make the final projection.
Extreme Gradient Boosting (XGB)	Extreme Gradient Boosting is a popular implementation of the Gradient Boosting algorithm that is optimized for speed and performance. It uses a technique called "gradient boosting with regularization," which adds a penalty term to the loss function to reduce overfitting. XGB is known for its high performance, scalability, and ability to handle large and complex data sets.

IV. PROPOSED METHOD

The proposed method utilizes both URLs and the textual contents of webpages from the raw dataset. Three distinct feature sets, ULF (URL linguistic features), CNF (Character level ngram features), and WCF (Web content features), are created.

- ULF-URL linguistic features, which include the conventional URL features.
- CNF-URL character level ngram features, which tokenize the URL's domain name using character level ngram.
- WCF-Web content features include webpage text.

Three separate NLP vectorization algorithms are employed to vectorize CNF and WCF. Subsequently, ULF, CNF, and WCF are amalgamated to form the final dataset for training and testing. The process flow of the proposed method is shown in Fig. 2.

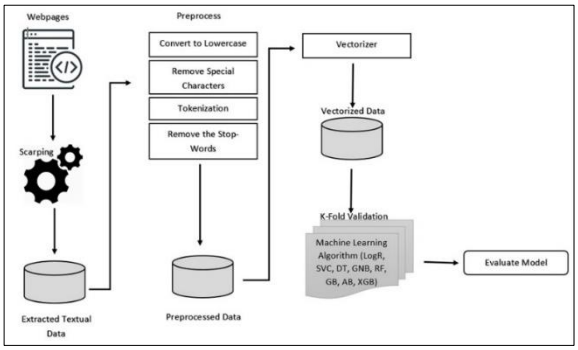


Fig. 2. Process flow.

TABLE III. URL FEATURES

No	Features	Description	No	Features	Description
1	ip_status	Presence of IP address in the URL	12	hyp_dom	Count the hyphens in domain name
2	dots_url	Count the dots in the URL.	13	at_dom	Count the @ in domain name.
3	slash_url	Count the / in the URL	14	underscr_dom	Count the underscores in the domain name
4	hyp_url	Count the hyphens in the URL	15	urllen	Length of the URL
5	hash_url	Count the # in the URL	16	num_url	Count the numbers in the URL
6	semi_url	Count the semicolons in the URL	17	alpha_url	Count the alphabet in the URL
7	and_url	Count the & in the URL	18	spl_url	Count the special symbols in the URL
8	underscr_url	Count the underscores in the URL	19	domlen	Length of domain name
9	http_url	Presence of http in the URL	20	num_dom	Count the numbers in the domain name
10	https_url	Presence of https in the URL	21	alpha_dom	Count the alphabet in the domain name
11	dots_dom	Count the dots in the domain name	22	spl_dom	Count the special symbols in the domain name
			23	url_class	Class of the URL either malicious (1)/benign (0)

A. Raw Dataset

The experiment's dataset, which contains both malicious and benign URLs, was gathered from the Kaggle URLs dataset [22]. The collection contains 450176 URLs. Classification is influenced by imbalanced datasets [23]. To prevent this problem, the experiment employs 6504 benign and 6478 malicious URLs.

B. Conventional Feature Extraction (URL Linguistic Features (ULF))

Conventional URL features include the dots in the URL, numbers in the URL, etc. Our experiments use only 23 lexical features (22 independent features and 1 dependent feature) of URLs, which are listed in Table III and saved as a separate file (conv fs.csv) as shown in Fig. 2.

C. Generating Character Level ngrams (Character level ngram features (CNF))

Character-level n-gram and word-level n-gram are two types of n-gram models used in natural language processing and machine learning. Both of these models are used for text analysis, but they operate at different levels of granularity [24]. To fully utilize the domain name of the URL, character level ngram is used in our experiment. A character-level n-gram model looks at sequences of characters in a text, regardless of the words they form. The model divides the text into n-grams, or consecutive groups of n characters called tokens, as shown in Fig. 3. Number of tokens for ngram is calculated by using the Eq. (3):

$$T_n = L - n + 1 \quad (3)$$

Where

- T_n=Total number of Tokens
- L=Length of the text
- n=Size of the ngram

For this experiment, the N value of the ngram is set between 3 and 7.

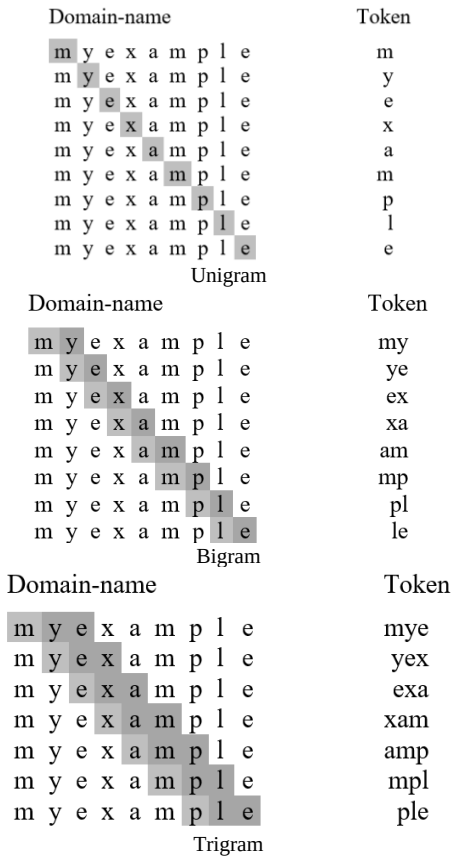


Fig. 3. Generating N-gram from the domain name.

D. Webpage Text Extraction

The process of extracting content from websites focuses on paragraph tags. The paragraph tag is a reasonable and straightforward way to organize the content of a web page because it is meant to denote a block of text that is unique from other blocks of text. After text extraction, text preprocessing begins. The preprocessing procedures are outlined in Algorithm 1.

E. Preprocess

The entire process of text vectorization is included in preprocessing. Text is cleansed by removing special characters and unnecessary components of the URL, such as the protocol, leaving only text and numbers. After text cleaning, it will be converted to lowercase, stop words will be removed, and lemmatized to reduce number of features during vectorization process. The procedures for preprocessing are summarized in Algorithm 2.

Algorithm 1: Text Extractions

```
Input: List of URLs (W)
Output: doc
Function Cont_Extraction (W, TagName)
Con=Connect(W)
Tag List = get Tag (Con, TagName)
For tag in TagList do
     doc = doc U getText(tag)
end for
                    return doc
```

The pseudo code of the method Cont_Extract () is used to extract textual contents from particular tags on the webpage. The Connect() function is used to establish the connection to the appropriate website. Following the establishment of the connection, the content of the given tag "TagName" is extracted and added to the string "doc". This "doc" will be used for preprocessing.

Algorithm 2: Data Preprocessing

```
Input: doc
Output: Corpus (T2 )
Function preProcess(doc)
Torg=read_Text(doc)
T1=clean_Text(Torg)
T1 = lower_Text(T1)
For token in T1 do
     if token is not in STOPWORDS of ENGLISH then
          T2 = T2 U Lemmatize(token)
     End if
end for
return T2
```

The pseudo-code of the preprocess() method is used to complete the preprocessing task for the given text document. The contents of the "doc" are read using the read_Text () function. The clean_Text() and lower_Text() routines convert text to uppercase and lowercase correspondingly. Stop words in a text are eliminated and lemmatized using the "for loop". The resultant corpus is to be vectorized.

F. Vectorization

Most machine learning algorithms take numeric feature vectors as input. Consequently, while working with text documents, required to convert each document into a numeric vector. This method is referred to as text vectorization. By employing various NLP techniques, such as count vectorizer, TFIDF, and hashing vectorizer, the generated tokens of text in the preprocessing and character level ngram are converted into a real-valued vector. The output of vectorization is a two-dimensional (2D) array. The vectorizer's features are set to 2500 and stored in a CSV file to limit the 2D array's size. As illustrated in Fig. 2, two files are created in our experiment, one for character level ngram tokens and another for textual content of the webpage (char level vec.csv & para text vec.csv).

G. Merging Files

After vectorization, three separate files (conv fs.csv, char level vec.csv & para text vec.csv) are combined into a single data file (combined.csv) as shown in Fig. 2. This file serves as a data file for the machine learning algorithm.

H. K-Fold Validation and Model Evaluation

The process uses a single parameter, K, to partition a data sample into k groups. When a particular number for k is selected, it may be used in place of k in the reference to the model, such as k=10 becoming 10-fold cross-validation [15]. In our experiment, a machine learning algorithm using 10-fold cross validation is fed the combined dataset (combined.csv). Four important metrics—precision, recall, accuracy, and f1-score—are used to assess the machine algorithm's performance.

V. Experimental Result

The experimental configuration consists of Windows 10, an I5 processor (3.2 GHz), and 8 GB of RAM. For programming, Python and sklearn package are utilized. Three distinct features, including URL Linguistic Features (ULF), Character-level Ngram Features (CNF), and Web Content Features (WCF), are generated for the experiments. Table II lists the seven most popular machine learning techniques used in the experiments. Each feature set and feature set combination (ULF+CNF+WCF) is examined independently for performance evaluation. Three different vectorizers were used to generate the features. A range of features between 250 and 2000 was taken for each trial. This range is known as "feature base."

1) URL linguistic feature (ULF): The features are retrieved from the URL alone by counting some characters in the URL and checking for the presence of the required pattern or characters in the URL. Table III provides the 22 features extracted from the URL. These are the most common features seen in the majority of existing research works. Fig. 2 shows the conventional feature extraction module extracting features from the Raw dataset and preparing a 2D array of values where rows represent URLs and columns represent the 22 features. The last column is the dependent feature called "class" which shows if the URL is benign (0) or malicious (1). The experiment's outcomes reveal that the random forest

achieves an accuracy of 98.31%, as depicted in both Table IV and Fig. 4. This outcome distinctly underscores the appropriateness of the selected features for the experiment.

TABLE IV. PERFORMANCE OF ULF

Type	Machine Learning Algorithm	Accuracy	Precision	Recall	F1-Score
URL Linguistic Features	LogR	95.06	99.48	90.58	94.82
	KNN	96.80	97.67	95.91	96.77
	GNB	91.19	99.44	82.82	90.36
	DT	96.66	95.76	97.84	96.74
	RF	**98.31**	**98.74**	**97.90**	**98.31**
	GB	96.83	99.06	94.55	96.74
	XGB	97.03	96.99	97.39	97.12

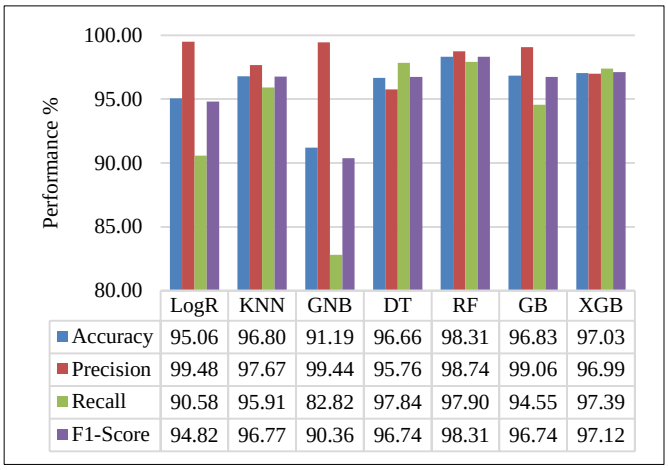


Fig. 4. Performance of ULF.

2) Character level-n-gram features (CNF): Character-level ngrams are generated by breaking a text into sequences of characters of a predetermined length. Character-level ngrams are useful for identifying patterns in text, specifically URL processing. In this experiment, character level ngram processing simply takes the URL's domain name into account. Character-level ngrams produce tokens. Following the generation of tokens, the tokens are vectorized using three distinct vectorizers for feature generation Experiments use seven machine learning algorithms to test the generated features and evaluate the performance. Tables V to VII show the outcomes of the various trials for the count, TFIDF, and hashing vectorizer, respectively. The results showed that using a 2000 feature base, count vectorizer + random forest achieves an accuracy of 90.87%, TFIDF vectorizer + random forest achieves an accuracy of 90.04%, and Hashing vectorizer + random forest achieves an accuracy of 92.95%. Fig. 5 depicts the performance comparison of three different vectorizers.

TABLE V. PERFORMANCE OF CNF WITH COUNT VECTORIZER

Feature Base	Machine Learning Algorithm	Accuracy	Precision	Recall	F1-Score
250	LogR	70.11	66.55	92.84	76.65
	KNN	68.88	70.82	69.28	69.11
	GNB	67.32	63.47	99.35	76.50
	DT	72.78	68.72	92.08	78.00
	RF	72.38	68.55	91.94	77.70
	GB	70.65	66.64	94.38	77.28
	XGB	71.72	67.86	92.37	77.48
500	LogR	71.78	70.42	82.53	75.16
	KNN	76.65	75.60	82.19	78.28
	GNB	68.77	64.59	98.83	77.20
	DT	81.12	78.45	88.75	82.77
	RF	81.80	78.75	89.69	83.40
	GB	70.80	67.25	92.62	77.01
	XGB	76.27	73.45	87.47	79.23
1000	LogR	77.57	76.19	83.78	79.29
	KNN	77.33	88.71	62.86	73.51
	GNB	67.97	72.96	61.44	65.96
	DT	86.37	83.12	93.27	87.59
	RF	88.71	85.82	93.56	89.37
	GB	70.98	67.27	93.32	77.28
	XGB	78.37	77.64	84.59	80.17
1500	LogR	79.56	77.78	85.66	81.09
	KNN	78.27	87.31	66.59	75.43
	GNB	66.75	81.46	44.30	57.13
	DT	87.97	84.40	94.72	89.00
	RF	89.93	87.04	94.69	90.54
	GB	71.38	67.43	93.92	77.63
	XGB	78.23	76.77	86.54	80.47
2000	LogR	81.35	79.29	87.11	82.66
	KNN	79.16	86.77	69.42	76.98
	GNB	68.40	83.38	46.62	59.58
	DT	88.97	85.70	94.81	89.81
	RF	**90.87**	**88.37**	**94.69**	**91.30**
	GB	71.48	67.43	94.04	77.70
	XGB	78.82	77.53	86.66	80.93

Configuring the count vectorizer with 2000 features results in an elevated accuracy of 90.87% and an F1-Score of 91.30% when applied in conjunction with the random forest algorithm. By incorporating n-grams, the count vectorizer not only captures individual words but also preserves contextual information regarding word combinations.

TABLE VI. PERFORMANCE OF CNF WITH TFIDF VECTORIZER

Feature Base	Machine Learning Algorithm	Accuracy	Precision	Recall	F1-Score
250	LogR	69.37	66.23	92.53	76.19
	KNN	69.84	69.60	75.97	71.87
	GNB	67.44	63.54	99.40	76.57
	DT	70.37	66.79	92.17	76.63
	RF	72.80	68.65	92.64	78.14
	GB	68.29	64.87	94.23	75.92
	XGB	72.62	68.57	92.76	78.10
500	LogR	71.38	70.61	80.73	74.41
	KNN	72.84	71.96	79.81	75.08
	GNB	68.87	64.67	98.64	77.20
	DT	78.71	75.75	88.79	81.11
	RF	79.54	76.41	89.61	81.87
	GB	72.43	68.32	93.35	78.07
	XGB	76.76	73.68	89.32	80.00
1000	LogR	76.57	74.83	84.05	78.64
	KNN	76.26	74.24	84.87	78.66
	GNB	69.97	73.48	66.84	69.21
	DT	85.18	81.71	93.32	86.74
	RF	85.68	82.14	94.09	87.29
	GB	72.96	69.63	91.23	77.90
	XGB	79.83	76.37	90.61	82.33
1500	LogR	77.62	75.61	85.61	79.74
	KNN	78.14	83.74	70.73	76.47
	GNB	67.86	82.42	46.14	58.90
	DT	87.41	84.21	94.29	88.62
	RF	87.09	83.48	94.92	88.46
	GB	72.02	69.52	89.67	77.08
	XGB	79.89	76.57	90.78	82.44
2000	LogR	78.49	76.05	86.92	80.59
	KNN	78.58	84.48	71.01	76.94
	GNB	69.49	84.35	48.27	61.18
	DT	86.54	83.12	94.75	88.10
	RF	90.04	87.01	95.17	90.72
	GB	69.44	67.31	88.18	74.98
	XGB	80.35	77.11	90.34	82.63

Configuring the TF-IDF vectorizer with 2000 features yields improved performance, achieving an accuracy of 90.04% and an F1-Score of 90.72% when coupled with the random forest algorithm. The incorporation of n-grams enhances the TF-IDF vectorizer, providing a more comprehensive representation of textual data.

TABLE VII. PERFORMANCE OF CNF WITH HASHING VECTORIZER

Feature Base	Machine Learning Algorithm	Accuracy	Precision	Recall	F1-Score
250	LogR	67.97	67.40	79.14	71.94
	KNN	66.61	62.76	99.41	76.01
	GNB	62.36	61.28	88.25	71.06
	DT	88.79	85.36	95.85	89.94
	RF	90.51	87.86	95.51	91.26
	GB	76.04	72.25	90.26	79.61
	XGB	88.49	85.04	94.81	89.43
500	LogR	71.01	69.73	82.11	74.61
	KNN	66.38	62.67	99.35	75.90
	GNB	65.35	63.39	90.45	73.31
	DT	88.28	84.76	95.45	89.45
	RF	90.84	88.34	95.48	91.51
	GB	76.80	73.01	89.73	79.96
	XGB	87.31	83.54	94.37	88.39
1000	LogR	77.38	75.37	84.93	79.33
	KNN	66.18	62.54	99.38	75.81
	GNB	65.78	63.47	92.34	74.04
	DT	90.34	87.56	95.45	91.08
	RF	92.45	91.03	94.98	92.80
	GB	77.16	73.99	88.45	79.97
	XGB	87.75	84.24	93.89	88.63
1500	LogR	79.51	76.81	87.53	81.40
	KNN	66.09	62.48	99.34	75.75
	GNB	67.15	64.00	93.33	74.90
	DT	90.56	87.50	95.83	91.27
	RF	91.92	89.92	95.11	92.30
	GB	77.82	74.07	90.37	80.77
	XGB	86.73	82.76	93.92	87.80
2000	LogR	80.27	76.94	89.27	82.23
	KNN	66.08	62.48	99.40	75.76
	GNB	70.83	67.06	94.27	77.33
	DT	90.93	88.10	95.91	91.60
	RF	**92.95**	**91.84**	**94.86**	**93.20**
	GB	78.00	74.35	89.80	80.78
	XGB	87.31	83.80	93.62	88.23

Configuring the Hashing vectorizer with 2000 features yields an elevated accuracy of 92.95% and an F1-Score of 93.20% when coupled with the random forest algorithm. The incorporation of n-grams into the Hashing vectorizer enhances the representation of sequential word combinations, capturing contextual information and thereby improving the overall effectiveness of classification.

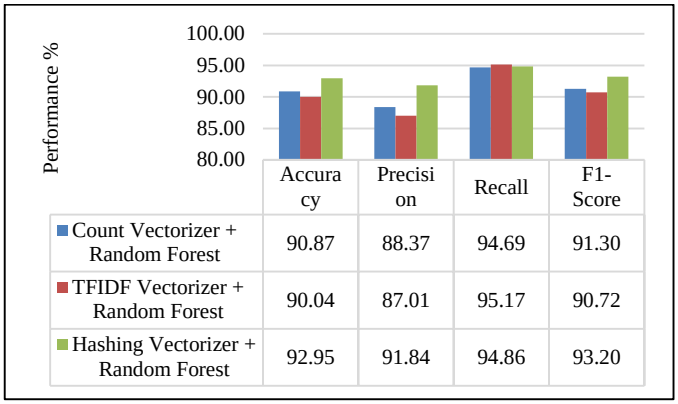


Fig. 5. Performance comparison of three vectorizers (2000 Features).

3) Web content features (WCF): The content of a webpage is the most informative option for analysis. Our experiment uses Beautiful Soup and Request in Python to analyze paragraph tag text. Text contents were preprocessed to get rid of stop words, special characters, etc. After that split the words into sentences to produce a corpus. This corpus serves as the input for the vectorizer, which produces features. To examine the generated features and assess performance, experiments utilize seven machine learning techniques. The results of the various trials for count, TFIDF, and hashing vectorizer, are displayed in Tables VIII to X. A performance comparison of three different vectorizers is shown in Fig. 6.

TABLE VIII. PERFORMANCE OF WCF WITH COUNT VECTORIZER

Feature Base	Machine Learning Algorithm	Accuracy	Precision	Recall	F1-Score
250	LogR	77.28	72.72	91.60	80.58
	KNN	78.35	74.87	88.90	80.76
	GNB	69.36	64.26	98.21	77.02
	DT	87.69	82.71	97.04	89.04
	RF	90.71	86.94	97.10	91.53
	GB	78.43	74.47	91.23	81.41
	XGB	85.82	81.37	95.20	87.39
500	LogR	79.58	75.04	91.83	82.16
	KNN	80.21	76.47	90.03	82.26
	GNB	70.89	65.51	98.18	77.92
	DT	88.91	84.55	97.30	90.16
	RF	91.54	88.22	97.05	92.23
	GB	79.20	75.15	91.45	81.96
	XGB	86.37	82.07	95.28	87.84
1000	LogR	82.98	78.65	93.13	84.89
	KNN	80.54	75.54	93.27	83.10
	GNB	73.36	67.53	98.26	79.41
	DT	90.17	86.00	96.85	90.94
	RF	93.15	90.27	97.18	93.50
	GB	79.53	75.46	91.62	82.22
	XGB	88.41	83.86	96.33	89.45
1500	LogR	84.13	79.64	93.86	85.84
	KNN	79.78	74.94	92.64	82.43
	GNB	73.77	67.82	98.50	79.71
	DT	90.73	86.51	97.41	91.47
	RF	93.58	90.97	97.19	93.88
	GB	79.58	75.57	91.71	82.30
	XGB	88.75	84.25	96.42	89.73
2000	LogR	85.76	81.11	94.91	87.21
	KNN	80.06	75.33	92.48	82.61
	GNB	75.31	69.08	98.36	80.60
	DT	92.44	88.99	97.39	92.90
	RF	**94.01**	**91.58**	**97.36**	**94.29**
	GB	79.58	75.73	91.28	82.22
	XGB	88.85	84.40	96.34	89.79

TABLE IX. PERFORMANCE OF WCF WITH TFIDF VECTORIZER

Feature Base	Machine Learning Algorithm	Accuracy	Precision	Recall	F1-Score
250	LogR	78.28	75.63	87.62	80.64
	KNN	77.98	73.51	91.94	81.18
	GNB	73.61	68.36	95.49	79.06
	DT	87.24	82.38	96.74	88.69
	RF	89.75	85.86	96.68	90.70
	GB	79.83	76.38	90.40	82.25
	XGB	86.04	81.87	94.74	87.50
500	LogR	80.89	78.40	88.58	82.68
	KNN	76.03	70.79	93.76	80.18
	GNB	77.05	71.43	95.45	81.19
	DT	88.29	83.51	97.22	89.57
	RF	91.78	88.68	96.96	92.44
	GB	80.80	77.42	90.60	82.97
	XGB	80.89	78.40	88.58	82.68
1000	LogR	83.34	80.96	89.58	84.65
	KNN	84.76	89.55	79.10	83.87
	GNB	77.60	71.70	96.71	81.81
	DT	90.46	86.08	97.31	91.20
	RF	93.27	90.52	97.10	93.60
	GB	80.18	76.59	91.11	82.65
	XGB	88.41	83.86	96.33	89.45
1500	LogR	85.03	82.53	90.65	86.07
	KNN	81.26	80.91	83.31	81.82
	GNB	79.19	73.32	97.16	83.00
	DT	92.42	89.07	97.19	92.86
	RF	93.70	91.12	97.25	94.00

	GB	80.75	77.16	91.37	83.11
	XGB	88.75	84.25	96.42	89.73
2000	LogR	86.13	83.92	91.02	87.01
	KNN	81.64	81.44	83.39	82.13
	GNB	81.59	75.71	97.04	84.59
	DT	92.20	88.62	97.31	92.67
	RF	**93.95**	**91.59**	**97.22**	**94.23**
	GB	79.54	75.69	91.31	82.21
	XGB	88.85	84.40	96.34	89.79

TABLE X. PERFORMANCE OF WCF WITH HASHING VECTORIZER

Feature Base	Machine Learning Algorithm	Accuracy	Precision	Recall	F1-Score
250	LogR	76.18	75.93	80.69	77.65
	KNN	83.65	80.93	89.23	84.67
	GNB	75.37	77.30	73.99	75.22
	DT	92.06	89.50	95.92	92.47
	RF	95.16	94.97	95.55	95.22
	GB	82.40	79.80	88.85	83.73
	XGB	91.53	88.74	95.80	92.00
500	LogR	81.94	81.21	84.75	82.62
	KNN	84.69	82.20	89.56	85.53
	GNB	78.55	81.39	75.36	77.95
	DT	92.66	90.22	96.37	93.05
	RF	94.92	94.21	95.94	95.02
	GB	81.81	78.89	89.53	83.44
	XGB	92.18	89.63	95.71	92.50
1000	LogR	85.39	83.96	88.67	86.01
	KNN	85.52	83.11	89.95	86.23
	GNB	79.35	82.89	75.42	78.65
	DT	91.64	88.60	96.47	92.20
	RF	94.98	93.87	96.53	95.11
	GB	80.81	77.38	90.20	82.84
	XGB	90.65	87.25	95.89	91.23
1500	LogR	86.17	84.56	89.43	86.72
	KNN	86.01	83.98	89.78	86.62
	GNB	82.32	86.06	77.83	81.50
	DT	91.57	88.61	96.33	92.13
	RF	**95.21**	**94.02**	**96.80**	**95.33**
	GB	80.71	77.14	90.61	82.88
	XGB	90.10	86.23	96.16	90.78
2000	LogR	86.83	85.21	89.98	87.34
	KNN	86.50	84.88	89.36	86.94
	GNB	83.42	88.21	77.54	82.40
	DT	92.41	89.67	96.60	92.86
	RF	94.97	93.62	96.84	95.13
	GB	80.97	77.44	90.71	83.07
	XGB	90.72	87.04	96.31	91.31

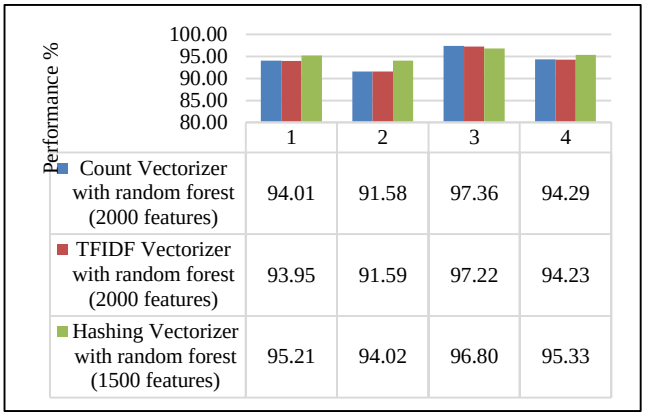


Fig. 6. Performance comparison of three vectorizers.

The findings demonstrated that when employing the 2000 feature base, the Count vectorizer with random forest achieves an accuracy of 94.01% and F1-Score of 94.29%. The TFIDF vectorizer with random forest reached an accuracy of 93.95% and an F1-score of 94.23%. However, the Hashing Vectorizer with Random Forest used a 1500 feature base and produced an accuracy of 95.21 % and an F1-score of 95.33%.

4) Combined features (CF): The combined feature set includes the features of ULF+CNF+WCF. Features are combined in row-wise. The main objective of combining feature sets is to reduce the feature set size and achieve higher accuracy. A range of features between 250 and 2000 was taken for each trial. So, the dataset for the trial is formed based on the following Eq. (4)

$$CF_n = ULF + CNF_n + WCF_n \quad (4)$$

Where n ϵ [250, 2000]

The features are merged in row-wise to create a single unified feature set. Tables XI to XIII show the outcomes of the various trials for the count, TFIDF, and hashing vectorizer, respectively. Fig. 7 depicts the performance comparison of three different vectorizers.

TABLE XI. PERFORMANCE OF CF WITH COUNT VECTORIZER

Feature Base	Machine Learning Algorithm	Accuracy	Precision	Recall	F1-Score
250	LogR	94.48	94.49	94.67	94.53
	KNN	86.59	82.70	93.66	87.64
	GNB	73.54	68.14	98.77	79.87
	DT	96.79	94.83	99.32	96.95
	RF	98.84	98.51	99.23	98.85
	GB	97.42	99.63	95.18	97.35
	XGB	98.51	98.09	99.03	98.54
500	LogR	94.94	93.92	96.39	95.07

	KNN	85.47	82.87	90.71	86.37
	GNB	76.14	70.30	98.61	81.40
	DT	98.24	97.23	99.37	98.27
	RF	**99.38**	**99.63**	**99.12**	**99.37**
	GB	97.81	99.83	95.79	97.76
	XGB	99.12	99.14	99.10	99.12
1000	LogR	96.03	94.34	98.15	96.15
	KNN	85.19	82.17	91.43	86.27
	GNB	78.51	72.52	98.44	82.88
	DT	98.64	97.91	99.44	98.66
	RF	**99.38**	**99.57**	**99.18**	**99.37**
	GB	97.94	99.90	95.97	97.90
	XGB	99.18	99.19	99.18	99.18
1500	LogR	96.17	94.59	98.10	96.28
	KNN	85.94	82.80	92.24	86.98
	GNB	79.17	73.08	98.63	83.34
	DT	98.24	97.19	99.44	98.28
	RF	99.33	99.63	99.03	99.33
	GB	97.91	99.89	95.92	97.86
	XGB	99.15	99.10	99.21	99.15
2000	LogR	96.47	94.89	98.38	96.57
	KNN	86.22	82.68	92.99	87.29
	GNB	79.96	73.92	98.53	83.87
	DT	98.68	97.96	99.44	98.69
	RF	99.23	99.46	99.00	99.23
	GB	97.93	99.87	95.97	97.88
	XGB	99.14	99.14	99.15	99.14

TABLE XII. PERFORMANCE OF CF WITH TFIDF VECTORIZER

Feature Base	Machine Learning Algorithm	Accuracy	Precision	Recall	F1-Score
250	LogR	95.76	96.84	94.66	95.72
	KNN	89.05	85.13	95.62	89.90
	GNB	73.40	67.83	99.75	79.97
	DT	95.65	93.79	99.24	96.17
	RF	99.27	99.38	99.15	99.27
	GB	97.48	99.75	95.20	97.42
	XGB	98.98	98.93	99.03	98.98
500	LogR	96.63	97.60	95.68	96.61
	KNN	89.79	88.37	92.03	90.06
	GNB	79.73	73.38	99.66	83.90
	DT	98.73	98.12	99.38	98.74
	RF	**99.33**	**99.54**	**99.12**	**99.33**
	GB	97.74	99.84	95.63	97.69
	XGB	99.07	99.08	99.06	99.07
1000	LogR	97.51	97.97	97.07	97.51
	KNN	86.96	83.04	94.75	88.23
	GNB	83.74	77.42	99.23	86.51
	DT	98.60	97.94	99.35	98.62
	RF	**99.35**	**99.48**	**99.21**	**99.34**
	GB	97.78	99.73	95.82	97.73
	XGB	99.26	99.30	99.23	99.26
1500	LogR	97.86	98.49	97.22	97.85
	KNN	87.98	84.95	95.52	89.42
	GNB	85.08	78.99	99.31	87.52
	DT	98.56	97.84	99.35	98.58
	RF	**99.38**	**99.68**	**99.07**	**99.37**
	GB	97.78	99.76	95.79	97.73
	XGB	98.98	98.93	99.03	98.98
2000	LogR	98.13	98.83	97.44	98.12
	KNN	78.09	73.13	97.48	82.73
	GNB	86.56	80.75	99.12	88.57
	DT	98.60	97.94	99.31	98.61
	RF	99.32	99.52	99.12	99.32
	GB	97.86	99.83	95.88	97.81
	XGB	99.18	99.13	99.23	99.18

TABLE XIII. PERFORMANCE OF CF WITH HASHING VECTORIZER

Feature Base	Machine Learning Algorithm	Accuracy	Precision	Recall	F1-Score
250	LogR	96.00	97.11	94.92	95.97
	KNN	91.71	90.58	93.32	91.87
	GNB	93.95	97.54	90.20	93.71
	DT	98.38	97.71	99.14	98.40
	RF	99.45	99.82	99.09	99.45
	GB	97.65	99.59	95.69	97.60
	XGB	99.41	99.63	99.20	99.41
500	LogR	97.19	97.87	96.53	97.18
	KNN	92.00	90.85	93.59	92.15
	GNB	91.93	90.71	93.89	92.16
	DT	98.26	97.43	99.18	98.28
	RF	99.15	99.29	99.03	99.15
	GB	97.68	99.87	95.48	97.62
	XGB	**99.50**	**99.68**	**99.32**	**99.50**
1000	LogR	97.57	98.43	96.71	97.55
	KNN	92.45	91.55	93.73	92.57
	GNB	90.58	87.13	96.30	91.28
	DT	98.57	97.96	99.23	98.58
	RF	99.33	99.68	98.98	99.33
	GB	97.80	99.79	95.79	97.75

	XGB	99.46	99.65	99.27	99.46
1500	LogR	98.17	99.25	97.07	98.14
	KNN	92.55	91.97	93.42	92.63
	GNB	90.27	86.22	96.70	90.99
	DT	98.52	97.75	99.37	98.54
	RF	99.29	99.54	99.04	99.29
	GB	97.97	99.87	96.06	97.93
	XGB	99.50	99.69	99.31	99.50
2000	LogR	97.91	98.71	97.11	97.90
	KNN	92.51	91.71	93.67	92.63
	GNB	91.97	88.63	96.90	92.46
	DT	98.44	97.63	99.35	98.47
	RF	99.03	99.06	99.03	99.03
	GB	97.86	99.70	96.00	97.81
	XGB	99.46	99.65	99.27	99.46

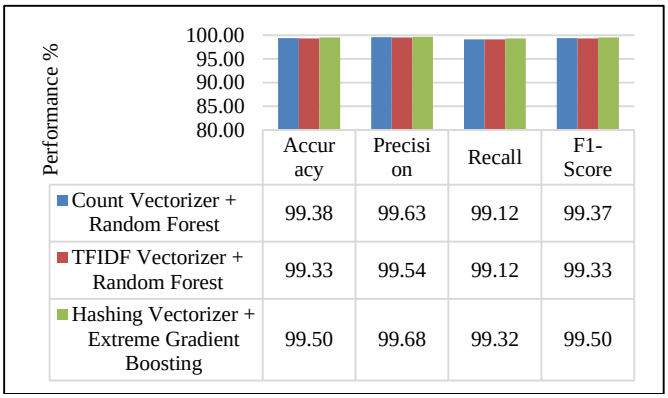


Fig. 7. Performance comparison of three vectorizers (500 Feature base).

The results showed that using a 500 features base, the Count vectorizer with random forest achieves an accuracy of 99.38% and an F1-score of 99.37%. The TFIDF vectorizer with random forest achieves an accuracy of 99.33% and an F1-score of 99.33% Hashing Vectorizer with extreme gradient boosting achieves an accuracy and F1-score of 99.50%.

All of the vectorizers employed in the experiment obtain higher accuracy and F1-score with 500 feature bases, especially the hashing vectorizer with extreme gradient boosting algorithm achieves the highest accuracy at 99.5%, as shown in Table XIV and Fig. 8. As compared to previous work, the proposed method significantly improves performance.

TABLE XIV. PERFORMANCE COMPARISON OF EXISTING AND PROPOSED METHOD

Author	Machine Learning Algorithm	Highest Accuracy
Saleem et al. [7]	Random Forest	93.0%
Saleem et al. [11]	Random Forest	99.0%
Malak et.al [12]	Naïve Bayes	96.0%
Kamel et.al [13]	Random Forest	95.51%
Lakshmanarao et.al[14]	Hashing vectorizer with Random Forest	97.50%
Joshi et.al [15]	Random Forest	92.0%
Hong et al. [16]	LA-FQ	94.16%
Josh et.al [17]	Random Forest	98.91%
Saleem et.al[18]	Weighted Soft Voting Classifier	99.10%
Ozgur et.al [19]	Random Forest	97.98%
Proposed Method	Hashing vectorizer with Extreme Gradient Boosting	99.50%

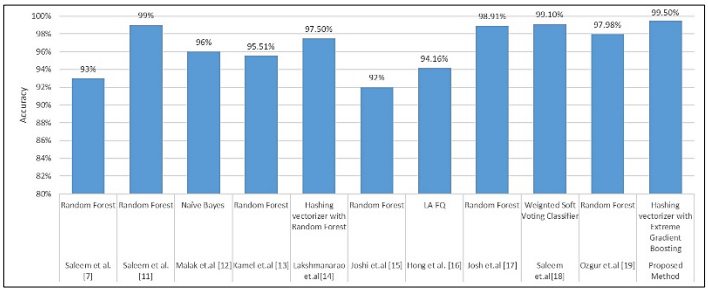


Fig. 8. Performance comparison of existing and proposed method.

VI. CONCLUSION

The majority of cybercrimes are committed using malicious links or malicious websites. Unintentionally visiting these websites or clicking on malicious links can have more serious effects, including the theft of private, sensitive information, security breaches, financial loss, and reputational damage. To find these kinds of dangerous websites on the internet, AI-based automated solutions are needed. For the detection method, this paper makes use of both URLs and web contents. Two different feature types, including ULF and CNF, are generated using URL of the webpage. Moreover, web page content is processed to generate features (WCF) for the detection procedure. Seven different machine learning methods are combined with three different vectorizers. Results of the study show that the proposed method, which combines an extreme gradient boosting algorithm with a hashing vectorizer, offers a better level of accuracy.

REFERENCES

[1] Data Breach Report. Internet: https://www.ibm.com/reports/data-breach. (Last access 18 oct 2024)

[2] Cybercrime magazine Internet: https://cybersecurityventures.com/cybercrime-damage-costs-10-trillion-by-2025/. (Last access 18 oct 2024)

[3] Hamadouche, Boudraa, Gasmi. Combining Lexical, Host, and Content-based features for Phishing Websites detection using Machine Learning Models. EAI Endorsed Transactions on Scalable Information Systems, vol.11, no.6, 2024. https://publications.eai.eu/index.php/sis/article/view/4421

[4] Saleem, Madhubala, Rajesh, Shaheetha, Arulkumar. Survey on Malicious URL Detection Techniques, 6th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 778-781, 2022. DOI: 10.1109/ICOEI53556.2022.9777221.

[5] MohammadMoein, Arash, Hardhik,Unveiling malicious DNS behavior profiling and generating benchmark dataset through application layer traffic analysis,Computers and Electrical Engineering, vol.118, 2024. https://doi.org/10.1016/j.compeleceng.2024.109436..

[6] Shivika, Feature-Rich Models and Feature Reduction for Malicious URLs Classification and Prediction, Iowa State University, 2019. https://dr.lib.iastate.edu/handle/20.500.12876/16711.

[7] Saleem, Sundarvadivazhagan, Vijayarangan, Veeramani., Malicious Webpage Classification Based on Web Content Features using Machine Learning and Deep Learning, International Conference on Green Energy, Computing and Sustainable Technology (GECOST), pp. 314-319, 2022. DOI: 10.1109/GECOST55694.2022.10010386.

[8] Rong, Yan, Jiefan, Binbin., Detection of malicious web pages based on hybrid analysis, Journal of Information Security and Applications, vol 35, pp. 68-74, 2017. (https://doi.org/10.1016/j.jisa.2017.05.008).

[9] Cho Do, Hoa Dinh, Tisenko, Malicious URL Detection based on Machine Learning, International Journal of Advanced Computer Science and Applications (IJACSA), vol. 11, no. 1, 2020. DOI: 10.14569/IJACSA.2020.0110119.

[10] Jiann, Yi, Kuan., Intelligent Visual Similarity-Based Phishing Websites Detection, Symmetry, 12, 1681. 2020. (https://doi.org/10.3390/sym12101681).

[11] Saleem, Vinodini, Kavitha, Lexical features based malicious URL detection using machine learning techniques, Materials Today: Proceedings, vol 47, 1, pp.163-166, 2021. https://doi.org/10.1016/j.matpr.2021.04.041.

[12] Malak, Fahd, Rami, Samiha, Dina, Hanan, Sara., An Assessment of Lexical, Network, and Content-Based Features for Detecting Malicious URLs Using Machine Learning and Deep Learning Models. Computational Intelligence and Neuroscience., 2022. DOI: 10.1155/2022/3241216.

[13] Kamel, Boukhalfa, Zakaria , Oussama., A new approach for the detection and analysis of phishing in social networks: the case of Twitter, Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS), pp. 1-8, 2020. DOI: 10.1109/SNAMS52053.2020.9336572.

[14] Lakshmanarao, Babu, Bala, Malicious URL Detection using NLP, Machine Learning and FLASK, International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), pp. 1-4, 2021. DOI: 10.1109/ICSES52305.2021.9633889.

[15] Saleem, Pradeepa, Justin, Madhubala, Hariraman, Vinodhini, SmishGuard: Leveraging Machine Learning and Natural Language Processing for Smishing Detection, International Journal of Advanced Computer Science and Applications, vol. 14, no. 11, 2023. DOI: 10.14569/IJACSA.2023.0141160.

[16] Hong, Zhaobin, Weijie, Xiangyan, Malicious Domain Names Detection Algorithm Based on Lexical Analysis and Feature Quantification, IEEE Access, vol. 7, pp. 128990-128999, 2019. DOI: 10.1109/ACCESS.2019.2940554.

[17] Jose, Ricardo , Emilio., Detection of algorithmically generated malicious domain names using masked N-grams, Expert Systems with Applications, vol.124, pp.156-163, 2019. https://doi.org/10.1016/j.eswa.2019.01.050.

[18] Saleem, Sundaravadivazhagan, Pradeepa, Justin, Karthikeyan., Weighted ensemble classifier for malicious link detection using natural language processing, International Journal of Pervasive Computing and Communications, 2023. https://doi.org/10.1108/IJPCC-09-2022-0312.

[19] Ozgur, Ebubekir, Onder, Banu., Machine learning based phishing detection from URLs, Expert Systems with Applications, vol.117, pp.345-357, 2019. https://doi.org/10.1016/j.eswa.2018.09.029.

[20] Benjamin, Rebecca, Tony., Applied Text Analysis with Python Enabling Language-Aware Data Products with Machine Learning, O'Reilly Media, Inc, 2018. ISBN-13: 978-1491963043.

[21] Aurélien Géron , Hands-On Machine Learning with Scikit-Learn and TensorFlow, O'Reilly Media, 2017. ISBN: 9781492032649.

[22] Malicious and Benign URLs: https://www.kaggle.com/datasets/siddharthkumar25/malicious-and-benign-urls.

[23] Jadhav, Mostafa, Elmannai, Khalid, An Empirical Assessment of Performance of Data Balancing Techniques in Classification Task, Applied Sciences, 12, 3928, 2022, https://doi.org/10.3390/app12083928.

[24] Li, Aletras, Improving Graph-Based Text Representations with Character and Word Level N-grams, arXiv:2210.05999. https://doi.org/10.48550/arXiv.2210.05999.

Enhanced IVIFN-ExpTODIM-MABAC Technique for Multi-Attribute Group Decision-Making Under Interval-Valued Intuitionistic Fuzzy Sets

Applications to College English Teaching Quality Evaluation

Bin Xie*
School of Architecture and Planning, Jiujiang University, Jiujiang, 332005, Jiangxi, China

***Abstract*—The evaluation of English teaching quality is crucial for enhancing teaching effectiveness. It helps teachers understand their teaching methods and students' learning outcomes through systematic assessment, thereby guiding teachers to adjust their teaching strategies. Additionally, the results of the evaluation provide decision-making support for educational management at schools, optimizing curriculum design and resource allocation. Regular evaluations of teaching quality motivate teachers for continuous professional development, improve teaching standards, and ensure that students achieve maximum growth and progress in their English learning journey. The assessment of college English teaching quality employs multi-attribute group decision-making (MAGDM). Techniques like Exponential TODIM (ExpTODIM) and MABAC are utilized to facilitate MAGDM. During the evaluation process, interval-valued intuitionistic fuzzy sets (IVIFSs) are utilized to handle fuzzy data. This research introduces a novel method, the interval-valued intuitionistic fuzzy number ExpTODIM-MABAC (IVIFN-ExpTODIM-MABAC) tailored for MAGDM under the framework of IVIFSs. To demonstrate its efficacy, a numerical example evaluating college English teaching quality is presented. Key contributions of this study include: (1) Extending the ExpTODIM-MABAC method to include IVIFSs with an Entropy model; (2) Utilizing Entropy to ascertain weights within IVIFSs; (3) Proposing the IVIFN-ExpTODIM-MABAC approach for MAGDM under IVIFSs; (4) Validating the approach with a numerical example and various comparative analyses of college English teaching quality.**

Keywords*—Multi-attribute group decision-making (MAGDM); interval-valued intuitionistic fuzzy sets (IVIFSs); ExpTODIM approach; MABAC approach; college English teaching quality evaluation

I. Introduction

The evaluation of English teaching quality is a critical component in educational management and pedagogical improvement. It not only focuses on the teaching skills and methods of teachers but also encompasses students' learning outcomes, the teaching environment, and curriculum content among other aspects [1]. The purpose and significance of evaluating English teaching quality are extensive and can be analyzed in the following dimensions: (1) Enhancing teaching quality. Teaching quality evaluation, through regular checks and feedback mechanisms, helps teachers recognize deficiencies in their teaching methods and techniques, thereby encouraging them to take measures for improvement. This evaluation process promotes continuous learning and self-improvement among teachers, ensuring that teaching methods stay current [2]. Additionally, teaching quality evaluations help teachers better manage course pacing and adjust teaching strategies to ensure the appropriateness and effectiveness of the content delivered. (2) Promoting student learning. Another important goal of teaching quality evaluation is to monitor and enhance students' learning outcomes [3-5]. During this process, teachers can understand students' interests, motivations, and challenges in learning English and then adopt personalized teaching approaches to meet the needs of different students [6]. For instance, for students with weaker foundations, teachers might increase tutoring time and use more interactive and visual aids to enhance their interest and efficiency in learning. (3) Optimizing curriculum design. Through the assessment of teaching quality, educational institutions can obtain direct feedback on which teaching contents and methods are most effective, which is crucial for curriculum design and textbook compilation [7, 8]. Evaluation results help education decision-makers understand changes in market and student needs, thereby optimizing the course structure, introducing new teaching contents, or eliminating outdated educational elements [9]. (4) Ensuring fairness and transparency in teaching. Regular teaching quality evaluations provide clear and fair assessment standards for all teachers and students. This standardized evaluation process helps eliminate biases and subjectivity, ensuring that every student is educated and assessed under the same standards, thus promoting fairness and transparency in educational activities [10, 11]. (5) Driving educational innovation. The results of teaching quality evaluations are a crucial basis for promoting educational innovation. Through evaluations, educational institutions can discover the effectiveness of new teaching methods and technologies, encouraging teachers to try and implement innovative teaching strategies. For example, with the development of information technology, many schools have begun experimenting with new teaching models such as blended learning and flipped classrooms. Teaching quality evaluations can effectively monitor the actual effects and adaptability of these new methods [12-14]. In conclusion, the evaluation of English teaching quality is a key tool for enhancing teaching effectiveness, ensuring educational fairness, and promoting comprehensive student development. It not only helps teachers and educational institutions understand the actual effects of

*Corresponding Author

teaching but also compels educational policies and practices to continuously adjust and improve. Therefore, establishing and maintaining a scientific and comprehensive teaching quality evaluation system is fundamental to improving educational quality and achieving educational modernization. Through such an evaluation system, we can better meet students' learning needs, enhance teachers' teaching abilities, and optimize the allocation of educational resources [15-17].

MAGDM is a collaborative decision-making process that brings together decision-makers from various backgrounds to evaluate multiple options across several attributes [18-21]. This method is prevalent in diverse sectors such as economics, management, engineering, and public policy, where complex decisions need balanced and comprehensive assessments. The essence of MAGDM lies in its ability to utilize the collective intelligence and resources of a group to pinpoint the most effective decision, considering a broad spectrum of factors. The purpose of MAGDM is multifaceted. Primarily, it aims to amalgamate diverse perspectives into a single decision-making process [22-25]. This integration is crucial in complex scenarios where no single decision-maker possesses enough information or expertise to cover all aspects of a decision comprehensively [26-28]. By involving multiple stakeholders, the decision-making process becomes more holistic and inclusive, thus enhancing the decision's quality and breadth. Moreover, MAGDM enhances the rationality and acceptance of decisions. Group discussions and evaluations increase transparency and fairness, fostering a decision-making environment where outcomes are more likely to be rational and broadly accepted. This is particularly important in settings where decisions have significant social, economic, or environmental impacts. Another critical aim of MAGDM is to optimize resource allocation. In environments where resources are scarce, this method facilitates a rational distribution based on thorough analysis and group consensus, ensuring that every decision maximizes the potential benefits from the available resources [29-32]. MAGDM also serves as a robust framework for dealing with complexity and uncertainty. By incorporating various data points and expert opinions, it provides a methodological approach to handle uncertainties inherent in dynamic environments, thereby stabilizing the decision-making process against external fluctuations [33-37]. The significance of MAGDM extends beyond just making effective decisions. It also enhances collaboration and communication among team members, who must share information and discuss options thoroughly before arriving at a consensus [38-42]. This interaction not only improves the decision outcomes but also strengthens team dynamics and cooperation. Furthermore, MAGDM promotes adaptability. As situations evolve and new information emerges, this decision-making framework can adapt, allowing decision-makers to revise their strategies promptly and efficiently [43-48]. This flexibility is vital in responding to changing conditions and maintaining the relevance and effectiveness of decisions. Lastly, the process of MAGDM increases both transparency and accountability. By documenting each decision-making step and its basis, the process ensures that all actions are traceable and decision-makers are accountable. This transparency is crucial for maintaining public trust, especially in decisions that affect broader communities or have significant public implications [38, 39, 49, 50]. In practical terms, MAGDM is seen in actions such as government policy formulation, corporate strategic planning, and new product development. For instance, in urban planning, government officials utilize MAGDM to integrate considerations like cost, public opinion, and environmental impact to devise comprehensive, sustainable urban transportation systems. In conclusion, MAGDM is more than just a decision-making tool; it is a critical process that taps into collective expertise to tackle complex problems efficiently and effectively. As the challenges in various fields grow more intricate, the role of MAGDM becomes increasingly vital, making it an indispensable element in the decision-making landscape of modern society. The evaluation of college English teaching quality involves MAGDM. IVIFSs [51] are used to represent fuzzy information during this evaluation. Various decision techniques have been proposed for MAGDM, managing the traditional ExpTODIM [52, 53] and MABAC [54, 55] techniques separately. However, there have been few or no techniques that combine the entropy model [56] with ExpTODIM [52, 53] and MABAC [54, 55] techniques under IVIFSs. Consequently, the IVIFN-ExpTODIM-MABAC technique is constructed to handle MAGDM under IVIFSs. A numerical example for evaluating college English teaching quality alongside different comparative analyses is provided to demonstrate the IVIFN-ExpTODIM-MABAC technique. The main objectives and motivations of this research include: (1) Enhancing the ExpTODIM-MABAC approach by integrating IVIFSs with the entropy framework. (2) Utilizing entropy to ascertain weights within IVIFSs. (3) Introducing the IVIFN-ExpTODIM-MABAC method for MAGDM in the context of IVIFSs. (4) Offering a numerical example and performing diverse comparative analyses to verify the IVIFN-ExpTODIM-MABAC model's effectiveness in assessing college English teaching quality.

The structure of the research is systematically organized. Section II introduces the IVIFSs. In Section III, the IVIFN-ExpTODIM-MABAC method is comprehensively proposed for MAGDM using IVIFSs. Section IV offers a numerical example that evaluates the quality of college English teaching, accompanied by comparative analyses. Section V concludes the paper with final remarks.

II. PRELIMINARIES

Atanassov [51] performed the IVIFSs.

Definition 1 [51]. The IVIFSs is introduced:

$$FI = \left\{ \left\langle FM\left(\zeta\right), FN\left(\zeta\right) \right\rangle \middle| \zeta \in X \right\} \tag{1}$$

where $FM\left(\zeta\right) \subset \left[0,1\right]$ is membership and $FN\left(\zeta\right) \subset \left[0,1\right]$ is non-membership, and $FM\left(\zeta\right), FN\left(\zeta\right)$ meet condition: $0 \le \sup FM\left(\zeta\right) + \sup FN\left(\zeta\right) \le 1$, $\forall \zeta \in X$. For convenience, $FI = \left(\left[FA, FB\right], \left[FC, FD\right]\right)$ is IVIFN.

Definition 2 [57]. Let $FI_1 = \left(\left[FA_1, FB_1\right], \left[FC_1, FD_1\right]\right)$ and $FI_2 = \left(\left[FA_2, FB_2\right], \left[FC_2, FD_2\right]\right)$ be IVIFNs, the operation laws are performed:

$$FI_1 \oplus FI_2 = \begin{pmatrix} \left[FA_1 + FA_2 - FA_1FA_2, FB_1 + FB_2 - FB_1FB_2\right], \\ \left[FC_1FC_2, FD_1FD_2\right] \end{pmatrix} \quad (2)$$

$$FI_1 \otimes FI_2 = \begin{pmatrix} \left[FA_1FA_2, FB_1FB_2\right], \\ \left[FC_1 + FC_2 - FC_1FC_2, FD_1 + FD_2 - FD_1FD_2\right] \end{pmatrix} \quad (3)$$

$$\lambda FI_1 = \left(\left[1-\left(1-FA_1\right)^{\lambda}, 1-\left(1-FB_1\right)^{\lambda}\right], \left[\left(FC_1\right)^{\lambda}, \left(FD_1\right)^{\lambda}\right]\right), \lambda > 0 \quad (4)$$

$$FI_1^{\lambda} = \left(\left[\left(FA_1\right)^{\lambda}, \left(FB_1\right)^{\lambda}\right], \left[1-\left(1-FC_1\right)^{\lambda}, 1-\left(1-FD_1\right)^{\lambda}\right]\right), \lambda > 0 \quad (5)$$

From Definition 2, several properties are established.

(1) $FI_1 \oplus FI_2 = FI_2 \oplus FI_1, FI_1 \otimes FI_2 = FI_2 \otimes FI_1, \left(\left(FI_1\right)^{\lambda_1}\right)^{\lambda_2} = \left(FI_1\right)^{\lambda_1\lambda_2}$;

(2) $\lambda\left(FI_1 \oplus FI_2\right) = \lambda FI_1 \oplus \lambda FI_2, \left(FI_1 \otimes FI_2\right)^{\lambda} = \left(FI_1\right)^{\lambda} \otimes \left(FI_2\right)^{\lambda}$;

(3) $\lambda_1 FI_1 \oplus \lambda_2 FI_1 = \left(\lambda_1 + \lambda_2\right) FI_1, \left(FI_1\right)^{\lambda_1} \otimes \left(FI_1\right)^{\lambda_2} = \left(FI_1\right)^{\left(\lambda_1 + \lambda_2\right)}$.

Definition 3 [58]. Let $FI_1 = \left(\left[FA_1, FB_1\right], \left[FC_1, FD_1\right]\right)$ and $FI_2 = \left(\left[FA_2, FB_2\right], \left[FC_2, FD_2\right]\right)$ be IVIFNs, both the score value (SV) and accuracy value (AV) of FI_1 and FI_2 are performed:

$$SV\left(FI_1\right) = \frac{FA_1 + FA_1\left(1-FA_1-FC_1\right) + FB_1 + FB_1\left(1-FB_1-FD_1\right)}{2} \quad (6)$$

$$SV\left(FI_2\right) = \frac{FA_2 + FA_2\left(1-FA_2-FC_2\right) + FB_2 + FB_2\left(1-FB_2-FD_2\right)}{2} \quad (7)$$

$$AV\left(FI_1\right) = \frac{FA_1 + FC_1 + FB_1 + FD_1}{2}, \quad (8)$$

$$AV\left(FI_2\right) = \frac{FA_2 + FC_2 + FB_2 + FD_2}{2} \quad (9)$$

For TI_1 and TI_2, from Definition 3, then

(1) *if* $SV\left(FI_1\right) < SV\left(FI_2\right)$, $FI_1 < FI_2$;

(2) *if* $SV\left(FI_1\right) > SV\left(FI_2\right), FI_1 > FI_2$;

(3) *if* $SV\left(FI_1\right) = SV\left(FI_2\right), AV\left(FI_1\right) < AV\left(FI_2\right)$, $FI_1 < FI_2$;

(4) *if* $SV\left(FI_1\right) = SV\left(FI_2\right), AV\left(FI_1\right) = AV\left(FI_2\right)$, $FI_1 = FI_2$.

Definition 4[59]. Let $FI_1 = \left(\left[FA_1, FB_1\right], \left[FC_1, FD_1\right]\right)$ and $FI_2 = \left(\left[FA_2, FB_2\right], \left[FC_2, FD_2\right]\right)$, the IVIFN Hamming distance (IVIFNHD) and IVIFN Euclidean distance (IVIFNED) are performed:

$$IVIFNHD\left(FI_1, FI_2\right) = \frac{1}{4}\begin{bmatrix} \left|FA_1 - FA_2\right| + \left|FB_1 - FB_2\right| \\ + \left|FC_1 - FC_2\right| + \left|FD_1 - FD_2\right| \end{bmatrix} \quad (10\text{-a})$$

$$IVIFNED\left(FI_1, FI_2\right) = \sqrt{\frac{1}{4}\begin{bmatrix} \left(FA_1 - FA_2\right)^2 + \left(FB_1 - FB_2\right)^2 \\ + \left(FC_1 - FC_2\right)^2 + \left(FD_1 - FD_2\right)^2 \end{bmatrix}} \quad (10\text{-b})$$

The IVIFWA and IVIFWG approach is performed [60].

Definition 5 [60]. Let $FI_j = \left(\left[FA_j, FB_j\right], \left[FC_j, FD_j\right]\right)\left(j = 1, 2, \cdots, n\right)$ be IVIFNs, the IVIFWA approach is performed:

$$IVIFWA_{f\omega}(FI_1, FI_2, \ldots, FI_n) = \bigoplus_{j=1}^{n}\left(f\omega_j FI_j\right) = \left(\left[1-\prod_{j=1}^{n}\left(1-FA_j\right)^{f\omega_j}, 1-\prod_{j=1}^{n}\left(1-FB_j\right)^{f\omega_j}\right], \left[\prod_{j=1}^{n}\left(FC_j\right)^{f\omega_j}, \prod_{j=1}^{n}\left(FD_j\right)^{f\omega_j}\right]\right) \quad (11)$$

where $f\omega=\left(f\omega_1, f\omega_2, ..., f\omega_n\right)^T$ be weight numbers of $FI_j=\left(\left[FA_j, FB_j\right], \left[FC_j, FD_j\right]\right)$, $f\omega_j>0, \sum_{j=1}^{n} f\omega_j = 1.$

Definition 6[61]. Let $FI_j=\left(\left[FA_j, FB_j\right], \left[FC_j, FD_j\right]\right)\left(j=1,2,\cdots,n\right)$ be IVIFNs, the IVIFWG approach is performed:

$$IVIFWG_{f\omega}\left(FI_1, FI_2, \ldots, FI_n\right) = \bigotimes_{j=1}^{n}\left(FI_j\right)^{f\omega_j} = \left(\left[\prod_{j=1}^{n}\left(FA_j\right)^{f\omega_j}, \prod_{j=1}^{n}\left(FB_j\right)^{f\omega_j}\right], \left[1-\prod_{j=1}^{n}\left(1-FC_j\right)^{f\omega_j}, 1-\prod_{j=1}^{n}\left(1-FD_j\right)^{f\omega_j}\right]\right) \quad (12)$$

where $f\omega=\left(f\omega_1, f\omega_2, ..., f\omega_n\right)^T$ be weight numbers of $FI_j=\left(\left[FA_j, FB_j\right], \left[FC_j, FD_j\right]\right)$, $f\omega_j>0, \sum_{j=1}^{n} f\omega_j = 1.$

III. IVIFN-ExpTODIM-MABAC Approach for MAGDM with Entropy

A. IVIFN-MAGDM Issues

Then IVIFN-ExpTODIM-MABAC approach is performed for MAGDM. Let $FA=\left\{FA_1, FA_2, \cdots, FA_m\right\}$ be alternatives and $FG=\left\{FG_1, FG_2, \cdots, FG_n\right\}$ be attributes with weight values $f\omega=\left(f\omega_1, f\omega_2, ..., f\omega_n\right)^T$, where $f\omega_j \in [0,1]$, $\sum_{j=1}^{n} f\omega_j = 1$ and experts $FE=\left\{FE_1, FE_2, \cdots, FE_q\right\}$ with weight values be $fw=\left\{fw_1, fw_2, \cdots, fw_q\right\}$. Then, IVIFN-ExpTODIM-MABAC technique is put forward for MAGDM.

Step 1. Perform the IVIFN-matrix $FR^{(t)}=\left[FR_{ij}^{(t)}\right]_{m\times n}=\left(\left[FA_{ij}^{(t)}, FB_{ij}^{(t)}\right], \left[FC_{ij}^{(t)}, FD_{ij}^{(t)}\right]\right)_{m\times n}$ and average matrix $FR=\left[FR_{ij}\right]_{m\times n}$:

$$FR^{(t)}=\left[FR_{ij}^{(t)}\right]_{m\times n}=\begin{array}{c} \\ FA_1 \\ FA_2 \\ \vdots \\ FA_m \end{array}\begin{array}{c} \begin{array}{cccc} FG_1 & FG_2 & \ldots & FG_n \end{array} \\ \begin{bmatrix} FR_{11}^{(t)} & FR_{12}^{(t)} & \ldots & FR_{1n}^{(t)} \\ FR_{21}^{(t)} & FR_{22}^{(t)} & \ldots & FR_{2n}^{(t)} \\ \vdots & \vdots & \vdots & \vdots \\ FR_{m1}^{(t)} & FR_{m2}^{(t)} & \ldots & FR_{mn}^{(t)} \end{bmatrix} \end{array} \quad (13)$$

$$FR=\left[FR_{ij}\right]_{m\times n}=\begin{array}{c} \\ FA_1 \\ FA_2 \\ \vdots \\ FA_m \end{array}\begin{array}{c} \begin{array}{cccc} FG_1 & FG_2 & \ldots & FG_n \end{array} \\ \begin{bmatrix} FR_{11} & FR_{12} & \ldots & FR_{1n} \\ FR_{21} & FR_{22} & \ldots & FR_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ FR_{m1} & FR_{m2} & \ldots & FR_{mn} \end{bmatrix} \end{array} \quad (14)$$

Based on IVIFWG technique, the $FR=\left[FR_{ij}\right]_{m\times n}=\left(\left[FA_{ij}, FB_{ij}\right], \left[FC_{ij}, FD_{ij}\right]\right)_{m\times n}$ is performed in Eq. (15):

$$FR_{ij}=\left(FR_{ij}^{(t)}\right)^{fw_1}\otimes\left(FR_{ij}^{(t)}\right)^{fw_2}\otimes\cdots\otimes\left(FR_{ij}^{(t)}\right)^{fw_q} = \left(\left[\prod_{t=1}^{q}\left(FA_{ij}^{(t)}\right)^{fw_t}, \prod_{t=1}^{q}\left(FB_{ij}^{(t)}\right)^{fw_t}\right], \left[1-\prod_{t=1}^{q}\left(1-FC_{ij}^{(t)}\right)^{fw_t}, 1-\prod_{t=1}^{q}\left(1-FD_{ij}^{(t)}\right)^{fw_t}\right]\right) \quad (15)$$

Step 2. The $FR=\left[FR_{ij}\right]_{m\times n}=\left(\left[FA_{ij}, FB_{ij}\right], \left[FC_{ij}, FD_{ij}\right]\right)_{m\times n}$ is normalized to $NFR=\left[NFR_{ij}\right]_{m\times n}=\left(\left[FA_{ij}^N, FB_{ij}^N\right], \left[FC_{ij}^N, FD_{ij}^N\right]\right)_{m\times n}$ in Eq. (16, 17).

For benefit attributes:

$$NFR_{ij}=\left(\left[FA_{ij}^N, FB_{ij}^N\right], \left[FC_{ij}^N, FD_{ij}^N\right]\right) = \left(\left[FA_{ij}, FB_{ij}\right], \left[FC_{ij}, FD_{ij}\right]\right) \quad (16)$$

For cost attributes:

$$NFR_{ij}=\left(\left[FA_{ij}^N, FB_{ij}^N\right], \left[FC_{ij}^N, FD_{ij}^N\right]\right) = \left(\left[FC_{ij}, FD_{ij}\right], \left[FA_{ij}, FB_{ij}\right]\right) \quad (17)$$

B. Perform the Attributes Weight through Entropy

Step 3. The weight numbers are important for MAGDM [62-64]. Entropy technique [56] is put forward weight numbers under IVIFSs. The normalized decision matrix is performed in Eq. (18):

$$NDM_{ij}=\frac{\begin{pmatrix}SV\left(\left[FA_{ij}^{N},FB_{ij}^{N}\right],\left[FC_{ij}^{N},FD_{ij}^{N}\right]\right)\\+AV\left(\left[FA_{ij}^{N},FB_{ij}^{N}\right],\left[FC_{ij}^{N},FD_{ij}^{N}\right]\right)+0.5\end{pmatrix}}{\sum_{i=1}^{m}\left(\begin{pmatrix}SV\left(\left[FA_{ij}^{N},FB_{ij}^{N}\right],\left[FC_{ij}^{N},FD_{ij}^{N}\right]\right)\\+AV\left(\left[FA_{ij}^{N},FB_{ij}^{N}\right],\left[FC_{ij}^{N},FD_{ij}^{N}\right]\right)+0.5\end{pmatrix}\right)},\tag{18}$$

The IVIFN Shannon entropy (IVIFNSE) is produced in Eq. (19):

$$IVIFNSE_{j}=-\frac{1}{\ln m}\sum_{i=1}^{m}NDM_{ij}\ln NDM_{ij}\tag{19}$$

and $NDM_{ij}\ln NDM_{ij}=0$ if $TIVIFNDM_{ij}=0$.

The weight numbers are performed in Eq. (20):

$$f\omega_{j}=\frac{1-IVIFNSE_{j}}{\sum_{j=1}^{n}\left(1-IVIFNSE_{j}\right)}\tag{20}$$

C. IVIFN-ExpTODIM-MABAC Approach for MAGDM

The IVIFN-ExpTODIM-MABAC is performed for MAGDM.

Step 4. Perform relative weight in Eq. (21):

$$rf\omega_{j}=f\omega_{j}\Big/\max_{j}f\omega_{j},\tag{21}$$

Step 5. The IVIFN dominance degree (IVIFNDD) is performed in Eq. (22) based on two kinds of distances measures:

$$IVIFNDD_{j}\left(FA_{i},FA_{i}\right)=\begin{cases}\frac{1}{2}\left(\begin{array}{l}\frac{rf\omega_{j}\times\left(1-10^{-\rho IVIFNED\left(NFR_{ij},NFR_{ij}\right)}\right)}{\sum_{j=1}^{n}rf\omega_{j}}\\+\frac{rf\omega_{j}\times\left(1-10^{-\rho IVIFNED\left(NFR_{ij},NFR_{ij}\right)}\right)}{\sum_{j=1}^{n}rf\omega_{j}}\end{array}\right) & \text{if } SV\left(NFR_{ij}\right)>SV\left(NFR_{ij}\right)\\ 0 & \text{if } SV\left(NFR_{ij}\right)=SV\left(NFR_{ij}\right)\\ \frac{1}{2}\left(\begin{array}{l}-\frac{1}{\theta}\frac{\sum_{j=1}^{n}rt\omega_{j}\times\left(1-10^{-\rho IVIFNED\left(NTR_{ij},NTR_{ij}\right)}\right)}{rt\omega_{j}}\\-\frac{1}{\theta}\frac{\sum_{j=1}^{n}rt\omega_{j}\times\left(1-10^{-\rho IVIFNED\left(NTR_{ij},NTR_{ij}\right)}\right)}{rt\omega_{j}}\end{array}\right) & \text{if } SV\left(NFR_{ij}\right)<SV\left(NFR_{ij}\right)\end{cases}\tag{22}$$

where θ is from [65] and $\rho\in[1,5]$ [66].

The IVIFNDD under FG_{j} is performed:

$$IVIFNDD_j\left(FA_i\right)=\left[IVIFNDD_j\left(FA_i,FA_t\right)\right]_{m\times m}$$

$$=\begin{array}{c} \\ FA_1 \\ FA_2 \\ \vdots \\ FA_m \end{array}\begin{array}{c} \begin{array}{cccc} FA_1 & FA_2 & \cdots & FA_m \end{array} \\ \begin{bmatrix} 0 & IVIFNDD_j\left(FA_1,FA_2\right) & \cdots & IVIFNDD_j\left(FA_1,FA_m\right) \\ IVIFNDD_j\left(FA_2,FA_1\right) & 0 & \cdots & IVIFNDD_j\left(FA_2,FA_m\right) \\ \vdots & \vdots & \cdots & \vdots \\ IVIFNDD_j\left(FA_m,FA_1\right) & IVIFNDD_j\left(FA_m,FA_2\right) & \cdots & 0 \end{bmatrix} \end{array}$$

(3) Perform the IVIFNDD of FA_i under FG_j :

$$IVIFNDD_j\left(FA_i\right)=\sum_{t=1}^{m}IVIFNDD_j\left(FA_i,FA_t\right) \tag{23}$$

The IVIFNDD is performed:

$$IVIFNDD=\left(IVIFNDD_{ij}\right)_{m\times n}$$

$$=\begin{bmatrix} & FG_1 & FG_2 & \cdots & FG_n \\ FA_1 & \sum_{t=1}^{m}IVIFNDD_1\left(FA_1,FA_t\right) & \sum_{t=1}^{m}IVIFNDD_2\left(FA_1,FA_t\right) & \cdots & \sum_{t=1}^{m}IVIFNDD_n\left(FA_1,FA_t\right) \\ FA_2 & \sum_{t=1}^{m}IVIFNDD_1\left(FA_2,FA_t\right) & \sum_{t=1}^{m}IVIFNDD_2\left(FA_2,FA_t\right) & \cdots & \sum_{t=1}^{m}IVIFNDD_n\left(FA_2,FA_t\right) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ FA_m & \sum_{t=1}^{m}IVIFNDD_1\left(FA_m,FA_t\right) & \sum_{t=1}^{m}IVIFNDD_2\left(FA_m,FA_t\right) & \cdots & \sum_{t=1}^{m}IVIFNDD_n\left(FA_m,FA_t\right) \end{bmatrix}$$

Step 6. Perform the IVIFNBAA (IVIFN Border Approximation Area).

$$IVIFNPIA_j=\max_{j=1}^{n}IVIFNDD_{ij},$$

$$IVIFNNIA_j=\min_{j=1}^{n}IVIFNDD_{ij} \tag{24}$$

$$NIVIFNDD_{ij}=1+\frac{IVIFNDD_{ij}-IVIFNNIA_j}{IVIFNPIA_j-IVIFNNIA_j}, \tag{25}$$

$$IVIFNBAA=\left[IVIFNBAA_j\right]_{1\times n} \tag{26}$$

$$IVIFNBAA_j=\prod_{j=1}^{n}\left(NIVIFNDD_{ij}\right)^{1/m}$$

$$=\prod_{j=1}^{n}\left(1+\frac{IVIFNDD_{ij}-IVIFNNIA_j}{IVIFNPIA_j-IVIFNNIA_j}\right)^{1/m} \tag{27}$$

Step 7. Perform the IVIFNDM (IVIFN distance measures) from IVIFNBAA:

$$IVIFNDM_{ij}=NIVIFNDD_{ij}-IVIFNBAA_j \tag{28}$$

Step 8. Perform the IVIFNWDM (IVIFN weighted distance measures) from IVIFNBAA.

$$IVIFNWDM_i=\sum_{j=1}^{n}\left(fw_j\times IVIFNDM_{ij}\right)$$

$$=\sum_{j=1}^{n}\left(fw_j\times\left(NIVIFNDD_{ij}-IVIFNBAA_j\right)\right) \tag{29}$$

Step 9. From the IVIFNWDM, the largest IVIFNWDM is the optimal choice.

IV. Numerical Example and Comparative Analysis

A. Numerical Example for College English Teaching Quality Evaluation

The evaluation of English teaching quality is a systematic process aimed at assessing the effectiveness of teaching and student learning outcomes to ensure that teaching activities effectively enhance students' English abilities. This process involves multiple aspects, including curriculum content, teaching methods, student engagement, the efficiency of learning resource use, and the teaching environment. Firstly, an essential aspect of the evaluation is the appropriateness of the curriculum content. This includes the rationality of the syllabus, the clarity of course objectives, and whether the depth and breadth of the content meet students' learning needs. Effective course design should cover all necessary language skills, such as listening, speaking, reading, and writing, and progressively enhance students' abilities through appropriate difficulty

gradients. Secondly, the innovation and diversity of teaching methods are key to the evaluation. Teachers should use various teaching strategies according to the specific needs and learning styles of students, such as cooperative learning, project-based learning, and multimedia teaching, to enhance the interactivity and interest of learning. The teacher's teaching style and abilities, such as clear expression, effective classroom management, and the timeliness and appropriateness of feedback to students, are important indicators for evaluation. Student engagement is another crucial metric for assessing teaching quality. Students' interaction in class, enthusiasm for participating in discussions, the quality and attitude toward completing assignments, and their mastery of learning materials all reflect the effectiveness of teaching. Additionally, the evaluation should also focus on students' learning motivation and self-development, including their self-assessment abilities and the achievement of learning objectives. During the evaluation process, it is also essential to fully utilize various assessment tools and technologies, such as classroom observations, student evaluations, standardized tests, and performance displays. These tools can provide direct and indirect evidence of teaching effectiveness, helping teachers and educational administrators make corresponding adjustments and improvements. Finally, the support of teaching resources and the environment is also crucial for ensuring teaching quality. Schools should provide sufficient learning materials, modern teaching facilities, and an environment conducive to learning. Moreover, continuous professional development support for teachers and resource updates are key to enhancing teaching quality. In summary, the evaluation of English teaching quality is a multi-faceted assessment process that not only focuses on teachers' teaching behaviors and students' learning outcomes but also involves curriculum design, teaching methods, student engagement, and teaching resources and environments. Through this comprehensive evaluation system, teaching quality can be effectively enhanced, ensuring students achieve the best outcomes in their English learning. The college English teaching quality evaluation is MAGDM process. Five potential English colleges $FA_i\left(i=1,2,3,4,5\right)$ are assessed with four attributes (Table I):

Five potential English colleges $FA_i\left(i=1,2,3,4,5\right)$ are assessed through linguistic scales (see Table II [67]) under experts $FE_t\left(t=1,2,3\right)$ with expert's weight $fw=\left(1/3,1/3,1/3\right)$.

TABLE I. FOUR ATTRIBUTES FOR SUSTAINABLE DEVELOPMENT EVALUATION IN HIGHER EDUCATION MANAGEMENT

Attribute	Attribute description
Student Learning Outcomes-FG_1	This indicator focuses on the specific learning achievements students gain through English instruction. It can be assessed by examining improvements in students' language skills, exam scores, and performance in simulated applications. This indicator reflects whether teaching activities can effectively enhance students' English capabilities.
Cost-effectiveness of Teaching-FG_2	The cost-related indicator primarily evaluates the relationship between input and output. In English teaching, this includes calculating the input of teaching resources (such as textbook costs, equipment investment, teacher training expenses, etc.) and comparing it with teaching outcomes. For example, the cost-effectiveness of teaching can be evaluated by analyzing the average improvement in students' scores or progression in language skills per unit cost.
Teacher Performance-FG_3	Return on educational investment is a crucial indicator of a university's economic efficiency. By assessing the returns on educational investments, universities can optimize resource allocation, enhance education quality, and improve students' employability, thereby increasing competitiveness and social impact.
Student Engagement and Feedback-FG_4	Active student participation is one of the crucial factors for successful teaching. This indicator measures through assessing students' interaction in class, homework completion, the level of activity in class discussions, and students' feedback on the teaching content. Student satisfaction with the teaching process and their suggestions can be obtained through surveys, interviews, or observations, reflecting the attractiveness and effectiveness of teaching activities.

TABLE II. LINGUISTIC SCALES AND IVIFNS

Linguistic information scale	IVIFNs
Exceedingly Bad-FEB	<[0.05,0.10], [0.85, 0.90]>
Very Bad-FVB	<[0.10,0.15], [0.75, 0.85]>
Bad-FB	<[0.15,0.20], [0.60, 0.70]>
Medium-FM	<[0.50,0.50], [0.50, 0.50]>
Good-FG	<[0.60,0.70], [0.15, 0.20]>
Very Good-FVG	<[0.75,0.85], [0.10, 0.15]>
Exceedingly Good-FEG	<[0.85,0.90], [0.05, 0.10]>

The IVIFN-ExpTODIM-MABAC is performed for college English teaching quality evaluation. $FR = \left[FR_{ij}^{(t)} \right]_{5\times4} \left(t = 1,2,3 \right)$ (Tables III to V).

Step 1. Describe the IVIFN-matrix

TABLE III. EVALUATION VALUES FOR FE_1

	FG_1	FG_2	FG_3	FG_4
FA_1	FM	FB	FVG	FVB
FA_2	FVB	FVG	FG	FM
FA_3	FVG	FG	FM	FB
FA_4	FVB	FM	FG	FVG
FA_5	FM	FVB	FVB	FVG

TABLE IV. EVALUATION INFORMATION FOR FE_2

	FG_1	FG_2	FG_3	FG_4
FA_1	FG	FVB	FVB	FM
FA_2	FG	FM	FB	FVG
FA_3	FB	FG	FVG	FVB
FA_4	FM	FM	FG	FB
FA_5	FM	FG	FVG	FB

TABLE V. EVALUATION INFORMATION FOR FE_3

	FG_1	FG_2	FG_3	FG_4
FA_1	FVG	FG	FM	FB
FA_2	FVG	FB	FVB	FM
FA_3	FM	FVG	FB	FG
FA_4	FVB	FM	FVG	FG
FA_5	FVB	FVG	FM	FG

Then, employing IVIFWG approach, the $FR = \left[FR_{ij} \right]_{5\times4}$ is performed (see Table VI).

TABLE VI. THE $FR = \left[FR_{ij} \right]_{5\times4}$

Alternatives	FG_1	FG_2
FA_1	([0.2431, 0.3410], [0.4571, 0.5589])	([0.4301, 0.4899], [0.3170, 0.3689])
FA_2	([0.1305, 0.2297], [0.7680, 0.8664])	([0.1470, 0.2459], [0.6521, 0.7503])
FA_3	([0.2540, 0.3498], [0.5460, 0.6452])	([0.3286, 0.4275], [0.4721, 0.5709])
FA_4	([0.1012, 0.1990], [0.7009, 0.7991])	([0.2017, 0.2975], [0.6007, 0.6985])
FA_5	([0.2233, 0.3196], [0.5778, 0.6754])	([0.3034, 0.4012], [0.4954, 0.5932])
Alternatives	FG_3	FG_5
FA_1	([0.2021, 0.3049], [0.5813, 0.6810])	([0.2568, 0.3502], [0.5350, 0.6348])
FA_2	([0.3698, 0.4691], [0.4190, 0.5172])	([0.0945, 0.1907], [0.7082, 0.8056])
FA_3	([0.0593, 0.1551], [0.7435, 0.8407])	([0.1124, 0.2086], [0.6881, 0.7879])
FA_4	([0.3925, 0.4928], [0.4057, 0.5045])	([0.0032, 0.1070], [0.8003, 0.8965])
FA_5	([0.0117, 0.1095], [0.7889, 0.8867])	([0.3058, 0.4026], [0.4942, 0.5900])

Step 2. Normalize the $FR = \left[FR_{ij} \right]_{5\times4}$ to $NFR = \left[NFR_{ij} \right]_{5\times4}$ (Table VII).

TABLE VII. THE $NFR = \left[NFR_{ij} \right]_{5\times4}$

Alternatives	FG_1	FG_2
FA_1	([0.2431, 0.3410], [0.4571, 0.5589])	([0.3170, 0.3689], [0.4301, 0.4899])
FA_2	([0.1305, 0.2297], [0.7680, 0.8664])	([0.6521, 0.7503], [0.1470, 0.2459])
FA_3	([0.2540, 0.3498], [0.5460, 0.6452])	([0.4721, 0.5709], [0.3286, 0.4275])
FA_4	([0.1012, 0.1990], [0.7009, 0.7991])	([0.6007, 0.6985], [0.2017, 0.2975])
FA_5	([0.2233, 0.3196], [0.5778, 0.6754])	([0.4954, 0.5932], [0.3034, 0.4012])
Alternatives	FG_3	FG_5
FA_1	([0.2021, 0.3049], [0.5813, 0.6810])	([0.2568, 0.3502], [0.5350, 0.6348])
FA_2	([0.3698, 0.4691], [0.4190, 0.5172])	([0.0945, 0.1907], [0.7082, 0.8056])
FA_3	([0.0593, 0.1551], [0.7435, 0.8407])	([0.1124, 0.2086], [0.6881, 0.7879])
FA_4	([0.3925, 0.4928], [0.4057, 0.5045])	([0.0032, 0.1070], [0.8003, 0.8965])
FA_5	([0.0117, 0.1095], [0.7889, 0.8867])	([0.3058, 0.4026], [0.4942, 0.5900])

Step 3. Perform the weight numbers (Table VIII):

TABLE VIII. THE WEIGHT NUMBERS

	FG_1	FG_2	FG_3	FG_4
weight numbers	0.2599	0.3620	0.1912	0.1869

Step 4. Perform the relative weight numbers (Table IX):

TABLE IX. THE RELATIVE WEIGHT NUMBERS

	FG_1	FG_2	FG_3	FG_4
relative weight	0.7180	1.0000	0.5282	0.5163

Step 5. Perform the $IVIFNDD = \left(IVIFNDD_{ij} \right)_{5\times4}$ (see Table X):

TABLE X. THE $IVIFNDD = \left(IVIFNDD_{ij} \right)_{5\times4}$

	FG_1	FG_2	FG_3	FG_4
FA_1	1.6044	-2.5732	0.6424	-1.1950
FA_2	-0.9564	2.0414	-3.5905	0.9154
FA_3	-1.1400	-0.1256	-0.3628	1.6101
FA_4	1.1026	-0.7552	0.3485	0.5389
FA_5	-1.4750	-2.1363	-1.6255	-1.3453

Step 6. Perform the IVIFNBAA (see Tables XI to XIII).

TABLE XI. THE IVIFNPIA AND IVIFNNIA

	FG_1	FG_2	FG_3	FG_4
IVIFNPIA	1.6044	2.0414	0.6424	1.6101
IVIFNNIA	-1.4750	-2.5732	-3.5905	-1.3453

TABLE XII. THE NIVIFNDD

	FG_1	FG_2	FG_3	FG_4
FA_1	1.0000	1.2824	1.0000	1.2093
FA_2	1.3743	2.0000	2.0000	1.3730
FA_3	1.3112	1.5068	1.5253	2.0000
FA_4	1.0417	1.7260	1.6653	1.7196
FA_5	2.0000	1.0000	1.6899	1.0000

TABLE XIII. THE IVIFNBAA

	FG_1	FG_2	FG_3	FG_4
IVIFNBAA	1.3662	1.3611	1.5838	1.4345

Step 7. Perform the IVIFNDM (see Tables XIV and XV):

TABLE XIV. THE IVIFNDM

	FG_1	FG_2	FG_3	FG_4
FA_1	0.6338	-0.3611	0.4162	-0.3836
FA_2	-0.1978	0.6389	-0.5838	0.3305
FA_3	-0.2574	0.1694	0.1788	0.5655
FA_4	0.4708	0.0329	0.3468	0.2031
FA_5	-0.3662	-0.2664	-0.1195	-0.4345

TABLE XV. THE WEIGHTED IVIFNDM

	FG_1	FG_2	FG_3	FG_4
FA_1	0.1647	-0.1307	0.0796	-0.0717
FA_2	-0.0514	0.2313	-0.1116	0.0618
FA_3	-0.0669	0.0613	0.0342	0.1057
FA_4	0.1224	0.0119	0.0663	0.0380
FA_5	-0.0952	-0.0964	-0.0229	-0.0812

Step 8. Perform the IVIFNWDM (see Table XVI).

TABLE XVI. THE IVIFNWDM

	IVIFNWDM	Order
FA_1	0.0419	4
FA_2	0.1300	3
FA_3	0.1343	2
FA_4	0.2385	1
FA_5	-0.2957	5

Step 9. From the IVIFNWDIM, the order is performed: $FA_4 > FA_3 > FA_2 > FA_1 > FA_5$ and TA_4 is the best English college.

B. Comparative Analysis

Then, the IVIFN-ExpTODIM-MABAC approach is compared with IVIFWA approach [60], IVIFWG approach [61], IVIFPWA approach [68], IVIFPWG approach [68], IVIFN-MABAC approach [69], IVIFN-Taxonomy approach [70] and IVIFN-TODIM approach [71]. The comparative results are performed in Table XVII.

TABLE XVII. ORDER FOR DIFFERENT APPROACHES

Approaches	Order
IVIFWA approach [60]	$FA_4 > FA_3 > FA_2 > FA_1 > FA_5$
IVIFWG approach [61]	$FA_4 > FA_3 > FA_1 > FA_2 > FA_5$
IVIFPWA approach [68]	$FA_4 > FA_3 > FA_2 > FA_1 > FA_5$
IVIFPWG approach [68],	$FA_4 > FA_3 > FA_1 > FA_2 > FA_5$
IVIFN-MABAC approach [69]	$FA_4 > FA_3 > FA_2 > FA_1 > FA_5$
IVIFN-Taxonomy approach [70]	$FA_4 > FA_3 > FA_2 > FA_1 > FA_5$
IVIFN-TODIM approach [71]	$FA_4 > FA_3 > FA_2 > FA_1 > FA_5$
IVIFN-ExpTODIM-MABAC approach	$FA_4 > FA_3 > FA_2 > FA_1 > FA_5$

In light with RW coefficients [72], the RW coefficient between IVIFWA approach [60], IVIFWG approach [61], IVIFPWA approach [68], IVIFPWG approach [68], IVIFN-MABAC approach [69], IVIFN-Taxonomy approach [70], IVIFN-TODIM approach [71] and IVIFN-ExpTODIM-MABAC approach is 1.0000, 0.9246, 1.0000, 0. 9246, 1.0000, 1.0000, 1.0000. This verifies the IVIFN-ExpTODIM-MABAC approach is effective.

V. CONCLUSION

The purpose of evaluating the quality of English teaching is to ensure the effectiveness of teaching methods, content, and resources, and to improve the teaching skills of educators and the learning outcomes of students. Through evaluation, strengths and weaknesses in teaching can be identified, promoting the professional development of teachers and innovation in teaching methods. Moreover, this process helps in designing courses that better meet the needs of students, enhancing the overall level of teaching, and ultimately aiming to improve students' English proficiency. The evaluation of college English teaching quality involves MAGDM. Currently, the TODIM-MABAC method is utilized in MAGDM contexts, with IVIFSs used to represent fuzzy decision data during the assessment. This research introduces the IVIFN-ExpTODIM-MABAC model for MAGDM, incorporating IVIFSs. To validate the effectiveness of the IVIFN-ExpTODIM-MABAC approach, a numerical example focusing on the evaluation of college English teaching quality is presented, alongside several comparative analyses. The key contributions of this study are: (1) The expansion of the ExpTODIM-MABAC method to include IVIFSs and an entropy model. (2) The use of entropy to assign weights within the context of IVIFSs. (3) The introduction of the IVIFN-ExpTODIM-MABAC method for MAGDM using IVIFSs. (4) The provision of a numerical example and execution of various comparative analyses to confirm the applicability of the IVIFN-ExpTODIM-MABAC model in assessing college English teaching quality.

REFERENCES

[1] Q. Deng, H. Yan, J. Sun, Research on the reform of college english teaching in the era of mooc, Agro Food Industry Hi-Tech, 28 (2017) 3388-3391.

[2] J. Liu, Innovation and development of english teaching modes in universities and colleges from internet perspective, Agro Food Industry Hi-Tech, 28 (2017) 1368-1370.

[3] Q.L. Xiong, College english mooc teaching on swot analysis, Educational Sciences-Theory & Practice, 18 (2018) 3529-3535.

[4] K. Gültekin, E. Mede, Blended teaching readiness of efl instructors and their perceptions about blended learning in english preparatory schools: A case from turkey, Journal of Research on Technology in Education, (2023) 17.

[5] Y.B. Zhang, Ieee, Application of blended teaching into the course of comprehensive english, in: 11th International Conference on Educational and Information Technology (ICEIT), Ieee, Sichuan Normal Univ, Chengdu, PEOPLES R CHINA, 2022, pp. 80-84.

[6] H. Wang, Exploration of the reform path of english teaching in higher vocational education under the background of mooc, Basic & Clinical Pharmacology & Toxicology, 127 (2020) 199-199.

[7] H.Y. Xie, Q. Mai, College english cross-cultural teaching based on cloud computing mooc platform and artificial intelligence, Journal of Intelligent & Fuzzy Systems, 40 (2021) 7335-7345.

[8] M.H. Du, Y.Q. Qian, Application of massive open online course to grammar teaching for english majors based on deep learning, Frontiers in Psychology, 12 (2022) 11.

[9] X.F. Li, X. Hui, An empirical study on mooc aided english listening teaching for english majors, Basic & Clinical Pharmacology & Toxicology, 128 (2021) 134-135.

[10] F.K. Li, X. Zhang, Artificial intelligence facial recognition and voice anomaly detection in the application of english mooc teaching system, Soft Computing, 27 (2023) 6855-6867.

[11] X.J. Xue, R.E. Dunham, Using a spoc-based flipped classroom instructional mode to teach english pronunciation, Computer Assisted Language Learning, 36 (2023) 1309-1337.

[12] L.Y. Zheng, K.C. Lee, Examining the effects of "small private online course and flipped-classroom"-based blended teaching strategy on first-year english-major students' achievements, Sustainability, 15 (2023) 26.

[13] L. Wei, An integrated decision-making framework for blended teaching quality evaluation in college english courses based on the double-valued neutrosophic sets, Journal of Intelligent & Fuzzy Systems, 45 (2023) 3259-3266.

[14] R.S. Wang, Economic benefits of blended teaching mode in the knowledge economy era: A case study of college english course, Journal of the Knowledge Economy, (2023) 48.

[15] Y. Qin, University blended english teaching based on big data analytics, Journal of Intelligent & Fuzzy Systems, 45 (2023) 9181-9197.

[16] Y.Z. Lei, A decision support method for designing the blended teaching effectiveness evaluation of english courses in universities based on interval-valued neutrosophic information, Journal of Intelligent & Fuzzy Systems, 45 (2023) 3267-3277.

[17] Z.J. Guo, An empirical analysis on the cultivation of english innovative

application ability of english majors in application-oriented universities based on dynamic optimization algorithm in blended teaching environment, Soft Computing, 27 (2023) 10839-10850.

[18] R. Verma, E. Alvarez-Miranda, Group decision-making method based on advanced aggregation operators with entropy and divergence measures under 2-tuple linguistic pythagorean fuzzy environment, Expert Systems with Applications, 231 (2023) 32.

[19] H.Y. Zhang, G.W. Wei, Location selection of electric vehicles charging stations by using the spherical fuzzy cpt-cocoso and d-critic method, Computational & Applied Mathematics, 42 (2023) 35.

[20] H.L. Wang, T. Mahmood, K. Ullah, Improved cocoso method based on frank softmax aggregation operators for t-spherical fuzzy multiple attribute group decision-making, International Journal of Fuzzy Systems, 25 (2023) 1275-1310.

[21] H.Y. Zhang, G.W. Wei, X.D. Chen, Spherical fuzzy dombi power heronian mean aggregation operators for multiple attribute group decision-making, Computational & Applied Mathematics, 41 (2022) 54.

[22] A. Mondal, S.K. Roy, J.M. Zhan, A reliability-based consensus model and regret theory-based selection process for linguistic hesitant-z multi-attribute group decision making, Expert Systems with Applications, 228 (2023) 18.

[23] N.N. Liao, H. Gao, R. Lin, G.W. Wei, X.D. Chen, An extended edas approach based on cumulative prospect theory for multiple attributes group decision making with probabilistic hesitant fuzzy information, Artificial Intelligence Review, 56 (2023) 2971-3003.

[24] C. Jana, H. Garg, M. Pal, B. Sarkar, G.W. Wei, Mabac framework for logarithmic bipolar fuzzy multiple attribute group decision-making for supplier selection, Complex & Intelligent Systems, (2023) 16.

[25] A. Hussain, K. Ullah, M. Mubasher, T. Senapati, S. Moslem, Interval-valued pythagorean fuzzy information aggregation based on aczel-alsina operations and their application in multiple attribute decision making, Ieee Access, 11 (2023) 34575-34594.

[26] Y. Zhang, S. Abdullah, I. Ullah, F. Ghani, A new approach to neural network via double hierarchy linguistic information: Application in robot selection, Engineering Applications of Artificial Intelligence, 129 (2024) 107581.

[27] S. Abdullah, I. Ullah, F. Ghani, Heterogeneous wireless network selection using feed forward double hierarchy linguistic neural network, Artificial Intelligence Review, 57 (2024) 191.

[28] S. Abdullah, A.O. Almagrabi, I. Ullah, A new approach to artificial intelligent based three-way decision making and analyzing s-box image encryption using topsis method, 11 (2023) 1559.

[29] F. Lei, Q. Cai, H. Wang, G. Wei, Z. Mo, An integrated group decision-making framework for evaluating the urban fire emergency management capability based on pdhlwphm and pdhlwpghm operator, Journal of Intelligent & Fuzzy Systems, 46 (2024) 3713-3760.

[30] F. Lei, Q. Cai, N.N. Liao, G.W. Wei, Y. He, J. Wu, C. Wei, Todim-vikor method based on hybrid weighted distance under probabilistic uncertain linguistic information and its application in medical logistics center site selection, Soft Computing, 27 (2023) 8541-8559.

[31] H. Zhu, R.M. Rodríguez, J.B. Zhao, Group decision making based on dice similarity measure and 2-dimension linguistic fuzzy weighted arithmetic aggregation operator for 2-dimension linguistic fuzzy variables, Engineering Applications of Artificial Intelligence, 135 (2024) 10.

[32] M. Zhao, Y.J. Wang, X.Y. Meng, X.J. Gou, Selection strategy of uniform expert evaluation scale in group decision making, Journal of the Operational Research Society, 75 (2024) 1178-1192.

[33] S.H. Wang, X.H. Pan, L. Martínez, A. Moreno-Albarracín, A novel interval type-2 fuzzy consensus reaching process model and group decision-making method for renewable energy investment, Engineering Applications of Artificial Intelligence, 133 (2024) 14.

[34] S. Wang, F. Chiclana, J.L. Chang, Y.M. Xing, J. Wu, A minimum cost-maximum consensus jointly driven feedback mechanism under harmonious structure in social network group decision making, Expert Systems with Applications, 238 (2024) 13.

[35] N. Zhang, Y.F. Zhou, J. Liu, G.W. Wei, Vikor method for pythagorean hesitant fuzzy multi-attribute decision-making based on regret theory, Engineering Applications of Artificial Intelligence, 126 (2023) 8.

[36] H.Y. Zhang, H.J. Wang, G.W. Wei, Spherical fuzzy todim method for magdm integrating cumulative prospect theory and critic method and its application to commercial insurance selection, Artificial Intelligence Review, 56 (2023) 10275-10296.

[37] Z.Y. Wang, Q. Cai, G.W. Wei, Enhanced todim based on vikor method for multi-attribute decision making with type-2 neutrosophic number and applications to green supplier selection, Soft Computing, (2023) 15.

[38] P. Koundouri, G.I. Papayiannis, E.V. Petracou, A.N. Yannacopoulos, Consensus group decision making under model uncertainty with a view towards environmental policy making, Environmental & Resource Economics, 87 (2024) 1611-1649.

[39] A. Kanchana, D. Nagarajan, K. Jacob, Neutrosophic multiplicative preference relations based on consensus analysis and additive consistency in group decision making: A goal programming approach, Expert Systems with Applications, 238 (2024) 16.

[40] J.F. Chu, P. Shu, Y.C. Liu, Y.Y. Wang, Y.M. Wang, A trust relationship network-based consensus model in large-scale todim group decision-making, Kybernetes, (2024) 28.

[41] F. Bakhshi, M. Ashtiani, An approach for reaching consensus in large-scale group decision-making focusing on dimension reduction, Complex & Intelligent Systems, 10 (2024) 4223-4251.

[42] Y. Liu, W.J. Chang, X.F. Jia, A group consensus model for multiple attributes group decision making with interval belief distribution and interval distributed preference relation, Group Decision and Negotiation, 32 (2023) 701-727.

[43] M. Ushada, R. Amalia, F. Trapsilawati, N.A.S. Putro, Group preference decision-making for the implementation of industry 4.0 in food and beverage smes, Technology Analysis & Strategic Management, 36 (2024) 1960-1977.

[44] M. Touqeer, R. Umer, N.A. Mohammad, M. Salimi, A. Ahmadian, Signed distance-based approach for multiple criteria group decision-making with incomplete information using interval type-2 neutrosophic numbers, Multimedia Tools and Applications, 83 (2024) 8439-8466.

[45] C.M. Stolle, B. Gula, R.J. Yu, Y. Huang, The impact of diversity on group decision-making in the face of the free-rider problem, Judgment and Decision Making, 19 (2024) 22.

[46] B.Q. Ning, G.W. Wei, The cross-border e-commerce platform selection based on the probabilistic dual hesitant fuzzy generalized dice similarity measures, Demonstratio Mathematica, 56 (2023) 34.

[47] B.Q. Ning, H.J. Wang, G.W. Wei, C. Wei, Several similarity measures of probabilistic dual hesitant fuzzy sets and their applications to new energy vehicle charging station location, Alexandria Engineering Journal, 71 (2023) 371-385.

[48] B.Q. Ning, H.J. Wang, G.W. Wei, C. Wei, Probabilistic dual hesitant fuzzy magdm method based on generalized extended power average operator and its application to online teaching platform supplier selection, Engineering Applications of Artificial Intelligence, 125 (2023) 32.

[49] S. Kheybari, M.R. Mehrpour, P. Bauer, A. Ishizaka, How can risk-averse and risk-taking approaches be considered in a group multi-criteria decision-making problem?, Group Decision and Negotiation, 33 (2024) 883-909.

[50] M. Jamil, F. Afzal, A. Maqbool, S. Abdullah, A. Akgül, A. Bariq, Multiple attribute group decision making approach for selection of robot under induced bipolar neutrosophic aggregation operators, Complex & Intelligent Systems, 10 (2024) 2765-2779.

[51] K.T. Atanassov, Operators over interval-valued intuitionistic fuzzy sets, Fuzzy Sets and Systems, 64 (1994) 159-174.

[52] A.B. Leoneti, L.F. Autran Monteiro Gomes, A novel version of the todim method based on the exponential model of prospect theory: The exptodim method, European Journal of Operational Research, 295 (2021) 1042-1055.

[53] H. Sun, Z. Yang, Q. Cai, G.W. Wei, Z.W. Mo, An extended exp-todim method for multiple attribute decision making based on the z-wasserstein distance, Expert Systems with Applications, 214 (2023) 14.

[54] K. Karunanithi, C. Han, C.J. Lee, W.C. Shi, L. Duan, Y. Qian, Identification of a hemodynamic parameter for assessing treatment outcome of edas in moyamoya disease, Journal of Biomechanics, 48 (2015) 304-309.

[55] M. Keshavarz Ghorabaee, E.K. Zavadskas, L. Olfat, Z. Turskis, Multi-criteria inventory classification using a new method of evaluation based on distance from average solution (edas), Informatica, 26 (2015) 435-451.

[56] C.E. Shannon, A mathematical theory of communication, Bell System Technical Journal, 27 (1948) 379-423.

[57] Z.S. Xu, R.R. Yager, Some geometric aggregation operators based on intuitionistic fuzzy sets, International Journal of General Systems, 35 (2006) 417-433.

[58] H.W. Liu, G.J. Wang, Multi-criteria decision-making methods based on intuitionistic fuzzy sets, European Journal of Operational Research, 179 (2007) 220-233.

[59] E. Szmidt, J. Kacprzyk, Distances between intuitionistic fuzzy sets, Fuzzy Sets and Systems, 114 (2000) 505-518.

[60] Z.X. Su, G.P. Xia, M.Y. Chen, Some induced intuitionistic fuzzy aggregation operators applied to multi-attribute group decision making, International Journal of General Systems, 40 (2011) 805-835.

[61] Z.S. Xu, J. Chen, On geometric aggregation ove interval-valued intuitionistic fuzzy information, in: 4th International Conference on Fuzzy Systems and Knowledge Discovery, Ieee Computer Soc, Haikou, PEOPLES R CHINA, 2007, pp. 466-+.

[62] J.H. Kim, B.S. Ahn, The hierarchical vikor method with incomplete information: Supplier selection problem, Sustainability, 12 (2020) 15.

[63] M.S.A. Khan, F. Khan, J. Lemley, S. Abdullah, F. Hussain, Extended topsis method based on pythagorean cubic fuzzy multi-criteria decision making with incomplete weight information, Journal of Intelligent & Fuzzy Systems, 38 (2020) 2285-2296.

[64] P.D. Liu, W.Q. Liu, Multiple-attribute group decision-making method of linguistic q-rung orthopair fuzzy power muirhead mean operators based on entropy weight, International Journal of Intelligent Systems, 34 (2019) 1755-1794.

[65] A. Tversky, D. Kahneman, Prospect theory: An analysis of decision under risk, Econometrica, 47 (1979) 263-291.

[66] A.B. Leoneti, L. Gomes, A novel version of the todim method based on the exponential model of prospect theory: The exptodim method, European Journal of Operational Research, 295 (2021) 1042-1055.

[67] J. Wang, Q. Cai, H.J. Wang, G.W. Wei, N.N. Liao, An integrated decision-making methodology for green supplier selection based on the improved ivif-cpt-mabac method, Journal of Intelligent & Fuzzy Systems, 44 (2023) 8535-8560.

[68] Z.S. Xu, Approaches to multiple attribute group decision making based on intuitionistic fuzzy power aggregation operators, Knowledge-Based Systems, 24 (2011) 749-760.

[69] Y.X. Xue, J.X. You, X.D. Lai, H.C. Liu, An interval-valued intuitionistic fuzzy mabac approach for material selection with incomplete weight information, Applied Soft Computing, 38 (2016) 703-713.

[70] L. Xiao, G. Wei, Y. Guo, X. Chen, Taxonomy method for multiple attribute group decision making based on interval-valued intuitionistic fuzzy with entropy, Journal of Intelligent & Fuzzy Systems, 41 (2021) 7031-7045.

[71] J.P. Lu, C. Wei, Todim method for performance appraisal on social-integration-based rural reconstruction with interval-valued intuitionistic fuzzy information, Journal of Intelligent & Fuzzy Systems, 37 (2019) 1731-1740.

[72] W. Sałabun, J. Wątróbski, A. Shekhovtsov, Are mcda methods benchmarkable? A comparative study of topsis, vikor, copras, and promethee ii methods, Symmetry, 12 (2020) 1549.

CoCoSo Framework for Management Performance Evaluation of Teaching Services in Sports Colleges and Universities with Euclidean Distance and Logarithmic Distance

Feng Li[1]*, Yuefei Wen[2]
Sergeant Education and Management College, Changsha Aviation Vocational and Technical College, Changsha, 410124, Hunan, China[1]
Qingyuan CITIC Primary School, Tianxin District, Changsha, 410004, Hunan, China[2]

***Abstract*—Sports colleges are the highest educational level in China's higher education system to cultivate sports professionals. They shoulder the arduous task of cultivating sports talents with innovative spirit and practical ability, and contribute to the country's sports and education undertakings. Through the study of the service performance of the teaching management departments in sports colleges, it is beneficial for teaching management workers to establish the central position of teaching work more firmly in their thoughts and actions, transform their work style, enhance their ideological awareness of serving teaching work, serve teaching, teachers, and students, closely focus on teaching work to provide various services, strive to improve the level of management services, and make improving service levels and optimizing service quality an important part of improving the teaching management level in sports colleges. The management performance evaluation of teaching services in sports colleges and universities is regarded as the defined multiple-attribute decision-making (MADM). Recently, the CoCoSo and entropy technique was utilized to cope with MADM. The double-valued neutrosophic sets (DVNSs) are utilized as a technique for characterizing fuzzy information during the management performance evaluation of teaching services in sports colleges and universities. In this study, double-valued neutrosophic number CoCoSo (DVNN-CoCoSo) technique is administrated for MADM in light with DVNN Euclidean distance (DVNNED) and DVNN Logarithmic distance (DVNNLD). Finally, numerical example for management performance evaluation of teaching services in sports colleges and universities is put forward to show the DVNN-CoCoSo technique. The major contribution of this study is administrated: (1) DVNN-CoCoSo technique is administrated for MADM in light with DVNNED and DVNNLD; (2) The objective weights are considered through entropy technique; (3) numerical example for management performance evaluation of teaching services in sports colleges and universities and some comparative analysis are administrated to verify the DVNN-CoCoSo technique.**

Keywords*—Multiple-Attribute Decision-Making (MADM); Double-Valued Neutrosophic Sets (DVNSs); CoCoSo technique; management performance evaluation of teaching services

I. INTRODUCTION

Cultivating talents is the basic function and fundamental task of higher education institutions in China, and the central task of schools is often teaching work. Therefore, teaching management plays a particularly important role [1]. Sports colleges and universities have an important mission in developing the national sports industry, advancing China towards a sports powerhouse, building an innovative country, and enhancing the country's competitiveness in development [2-4]. Sports colleges have various functions such as cultivating senior specialized sports talents for national construction, conducting sports scientific research, and engaging in social services [5-7]. They actively adapt to the development needs of the national sports industry with sports as their characteristics [8-10]. Teaching management is the core function of the teaching management department, which has unique characteristics of complexity, complexity, and repetition. In the process of development, sports colleges have gradually shown a trend of expanding their scale, comprehensive professional settings, and diversified educational levels [11, 12]. The difficulty and complexity of teaching management are increasing, and with the continuous deepening of information technology in teaching management, the requirements for the quality and information acquisition ability of managers are becoming higher and higher [13, 14]. With the continuous deepening of the implementation of the credit system, the transition from the academic year system to academic year credit system, and the transformation to the credit system, the original teaching management work methods, management methods, and systems seem somewhat inadequate and difficult to adapt to. The number of disciplines and majors in sports colleges is increasing, and the number of students on campus is also increasing [15-18]. The unified management by the school appears to have a huge coverage, and it is also difficult to achieve effective management after implementing secondary college management. In the practice of teaching management, we may encounter such phenomena: the school's decision-making and work ideas cannot be well implemented. There are more discussions on teaching work than practical ones, and the problems raised during the meetings are often not solved [19, 20]. The same problems are raised year by year and occur year by year. The school's rules and regulations lack a certain degree of constraint ability; Inadequate work arrangement; The execution process is sloppy, etc. Although schools have attempted to link the implementation of their respective departments with performance, the results have been minimal.

*Corresponding Author.

Low efficiency in handling affairs, unreasonable management structure, low overall business quality of management personnel, poor channels for uploading and issuing information, and emphasis on form over connotation in goal management and management performance evaluation [21-24]. The management methods are not entirely scientific, the hierarchical and categorical distinctions in management are insufficient, the system is rigid, the response to new things is slow, and the phenomenon of habitual defense is obvious [25, 26]. Political management is too broad, administrative intervention in teaching management is too strong, the official centric ideology is severe, there are too many checkpoints, work transparency is poor, coordination is insufficient, the tendency towards government is obvious, and the rule of man is greater than the rule of law. These phenomena have been troubling many workers engaged in teaching management [27]. Sports colleges are the highest educational level in China's higher education system to cultivate sports professionals. They shoulder the arduous task of cultivating sports talents with innovative spirit and practical ability, and contribute to the country's sports and education undertakings [28-30]. The construction of teaching management performance is achieved by adopting scientific management methods, mobilizing the enthusiasm, initiative, and creativity of the majority of teaching staff in their work, correcting their service attitude, enhancing their consciousness of actively serving teaching work, thereby improving work efficiency and ability, and better implementing the school's policies and guidelines [31-33].

The management performance evaluation of teaching services in sports colleges and universities is regarded as the defined MADM. Recently, the CoCoSo [34] and entropy [35] has been used to cope with MADM. The DVNSs [36] are used as a technique for characterizing fuzzy information during the management performance evaluation of teaching services in sports colleges and universities. Furthermore, many techniques employed CoCoSo technique [34] and entropy [35] separately to manage the MADM. Until now, no or few techniques have been administrated on entropy technique [35] and CoCoSo [34] under DVNSs. Therefore, the DVNN-CoCoSo model is founded to manage the MADM. Finally, numerical example for management performance evaluation of teaching services in sports colleges and universities and comparative analysis is administrated to validate the DVNN-CoCoSo model. The major research motivation of this work is managed: (1) the novel MADM is put forward based on CoCoSo and entropy technique under DVNSs. (2) The objective weights are considered through entropy technique. (3) The new MADM technique based on DVNN-CoCoSo technique is proposed for management performance evaluation of teaching services in sports colleges and universities. (4) numerical example for management performance evaluation of teaching services in sports colleges and universities and comparative analysis are employed to prove the DVNN-CoCoSo model.

The framework of this study is administrated. Section II introduces the DVNSs. In Section III, the DVNN-CoCoSo technique is administrated for MADM. In Section IV, numerical example for blended teaching quality decision evaluation is administrated and comparative analysis is conducted. The final study ends in Section V.

II. PRELIMINARIES

Kandasamy [36] administrated the DVNSs.

Definition 1 [36]. The DVNSs RA in Θ is put forward:

$$RA=\left\{\left(\begin{matrix}\theta,RT_A(\theta),RIT_A(\theta),\\ RIF_A(\theta),RF_A(\theta)\end{matrix}\right)\middle|\theta\in\Theta\right\}. \tag{1}$$

where $RT_A(\theta)$ is truth-membership, $RIT_A(\theta)$ is indeterminacy leaning towards $RT_A(\theta)$, $RIF_A(\theta)$ is indeterminacy leaning towards $RF_A(\theta)$, $RF_A(\theta)$ is falsity-membership,

$RT_A(\theta),RIT_A(\theta),RIF_A(\theta),RF_A(\theta)\in[0,1]$,

$0\le RT_A(\theta)+RIT_A(\theta)+RIF_A(\theta)+RF_A(\theta)\le 4$.

The DVNN is expressed as $RA=(RT_A,RT_A,RIF_A,RF_A)$, where

$RT_A,RIT_A,RIF_A,RF_A\in[0,1]$

$0\le RT_A+RIT_A+RIF_A+RF_A\le 4$.

Definition 2 [36]. Let $RA=(RT_A,RT_A,RIF_A,RF_A)$ be the DVNN, the score value is administrated:

$$SV(RA)=\frac{(2+RT_A+RIT_A-RIF_A-RF_A)}{4},$$

$$SV(RA)\in[0,1] \tag{2}$$

Definition 3 [36]. Let $RA=(RT_A,RT_A,RIF_A,RF_A)$ be the DVNN, the accuracy value is administrated:

$$AV(RA)=\frac{(RT_A+RIT_A+RIF_A+RF_A)}{4},$$

$$AV(RA)\in[0,1] \tag{3}$$

The order for DVNNs is administrated.

Definition 4[36]. Let $RA=(RT_A,RT_A,RIF_A,RF_A)$ and $RB=(RT_B,RT_B,RIF_B,RF_B)$,

$$SV(RA)=\frac{(2+RT_A+RIT_A-RIF_A-RF_A)}{4},$$

$$SV(RB)=\frac{(2+RT_B+RIT_B-RIF_B-RF_B)}{4},$$

$AV\left(RA\right)=\frac{\left(RT_A+RIT_A+RIF_A+RF_A\right)}{4}$,

$AV\left(RB\right)=\frac{\left(RT_B+RIT_B+RIF_B+RF_B\right)}{4}$, if $SV\left(RA\right)<SV\left(RB\right)$, $RA<RB$; if $SV\left(RA\right)=SV\left(RB\right)$, (1)if $AV\left(RA\right)=AV\left(RB\right)$, $RA=RB$; (2) if $AV\left(RA\right)<AV\left(RB\right)$, $RA<RB$.

Definition 5 [36]. Let $RA=\left(RT_A,RT_A,RIF_A,RF_A\right)$ and $RB=\left(RT_B,RT_B,RIF_B,RF_B\right)$ be two DVNNs, the operations are administrated:

$$(1)\ RA\oplus RB=\left(RT_A+RT_B-RT_ART_B,RIT_A+RIT_B-RIT_ARIT_B,RIF_ARIF_B,RF_ARF_B\right);$$
$$(2)\ RA\otimes RB=\left(RT_ART_B,RIT_ARIT_B,RIF_A+RIF_B-RIF_ARIF_B,RF_A+RF_B-RF_ARF_B\right);$$
$$(3)\ \lambda RA=\left(1-\left(1-RT_A\right)^{\lambda},1-\left(1-RIT_A\right)^{\lambda},\left(RIF_A\right)^{\lambda},\left(RF_A\right)^{\lambda}\right),\lambda>0;$$
$$(4)\left(RA\right)^{\lambda}=\left(\left(RT_A\right)^{\lambda},\left(RIT_A\right)^{\lambda},1-\left(1-RIF_A\right)^{\lambda},1-\left(1-RF_A\right)^{\lambda}\right),\lambda>0.$$

Definition 6[36]. Let $RA=\left(RT_A,RT_A,RIF_A,RF_A\right)$ and $RB=\left(RT_B,RT_B,RIF_B,RF_B\right)$, the DVNN Euclidean distance (DVNNED) between $RA=\left(RT_A,RT_A,RIF_A,RF_A\right)$ and $RA=\left(RT_A,RT_A,RIF_A,RF_A\right)$ is:

$$ED\left(RA,RB\right)=\sqrt{\frac{1}{4}\left(\begin{array}{l}\left|RT_A-RT_B\right|^2+\left|RIT_A-RIT_B\right|^2\\+\left|RIF_A-RIF_B\right|^2+\left|RF_A-RF_B\right|^2\end{array}\right)} \quad (4)$$

Definition 7. Let $RA=\left(RT_A,RT_A,RIF_A,RF_A\right)$ and $RB=\left(RT_B,RT_B,RIF_B,RF_B\right)$, the DVNN Logarithmic distance (DVNNLD) between $RA=\left(RT_A,RT_A,RIF_A,RF_A\right)$ and $RB=\left(RT_B,RT_B,RIF_B,RF_B\right)$ is administrated:

$$DVNNLD\left(RA,RB\right)=\frac{1}{4}\left(\begin{array}{l}RT_A\log\frac{2RT_A}{RT_A+RT_B}+RT_B\log\frac{2RT_B}{RT_A+RT_B}\\+RIT_A\log\frac{2RIT_A}{RIT_A+RIT_B}+RIT_B\log\frac{2RIT_B}{RIT_A+RIT_B}\\+RIF_A\log\frac{2RIF_A}{RIF_A+RIF_B}+RIF_B\log\frac{2RIF_B}{RIF_A+RIF_B}\\+RF_A\log\frac{2RF_A}{RF_A+RF_B}+RF_B\log\frac{2RF_B}{RF_A+RF_B}\end{array}\right) \quad (5)$$

III. DVNN-CoCoSo Approach for MADM with Entropy Weight

The DVNN-CoCoSo technique is administrated for MADM. Let $RA=\left\{RA_1,RA_2,\cdots,RA_m\right\}$ be alternatives, $RG=\left\{RG_1,RG_2,\cdots,RG_n\right\}$ be attributes with weight rw, where $rw_j\in\left[0,1\right]$, $\sum_{j=1}^{n}rw_j=1$. Suppose that assessed information are DVNNs $RM=\left(RM_{ij}\right)_{m\times n}=\left(RT_{ij},RIT_{ij},RIF_{ij},RF_{ij}\right)_{m\times n}$.

Then, DVNN-CoCoSo technique is put forward MADM (see Fig. 1).

Step 1. Put forward the DVNN-matrix $RM=\left(RM_{ij}\right)_{m\times n}=\left(RT_{ij},RIT_{ij},RIF_{ij},RF_{ij}\right)_{m\times n}$

$$RM=\left[RM_{ij}\right]_{m\times n}=\begin{bmatrix}RM_{11}&RM_{12}&\cdots&RM_{1n}\\RM_{21}&RM_{22}&\cdots&RM_{2n}\\\vdots&\vdots&\vdots&\vdots\\RM_{m1}&RM_{m2}&\cdots&RM_{mn}\end{bmatrix} \quad (6)$$

$$RM_{ij}=\left(RT_{ij},RIT_{ij},RIT_{ij},RF_{ij}\right) \quad (7)$$

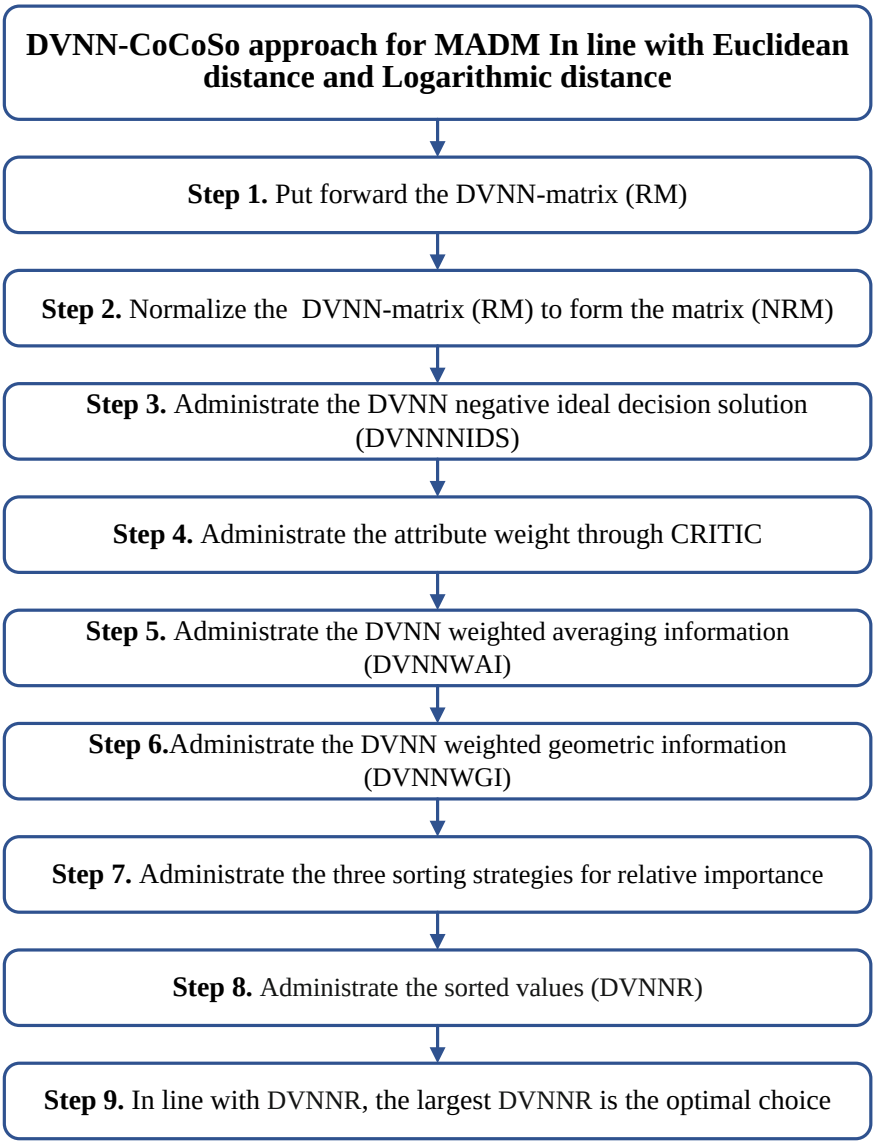


Fig. 1. DVNN-CoCoSo technique for MADM in line with Euclidean distance and Logarithmic distance.

Step 2. Normalize the

$$UR = \left(UR_{ij}\right)_{m\times n} = \left(UT_{ij}, UIT_{ij}, UIF_{ij}, UF_{ij}\right)_{m\times n}$$

into

$$NRM = \left(NRM_{ij}\right)_{m\times n} = \left(RT_{ij}^{N}, RIT_{ij}^{N}, RIF_{ij}^{N}, RF_{ij}^{N}\right)_{m\times n}$$

$$NRM_{ij} = \left(RT_{ij}^{N}, RIT_{ij}^{N}, RIF_{ij}^{N}, RF_{ij}^{N}\right) = \begin{cases} \left(RT_{ij}, RIT_{ij}, RIF_{ij}, RF_{ij}\right), & RG_j \text{ is benefit attribute} \\ \left(RF_{ij}, RIF_{ij}, RIT_{ij}, RT_{ij}\right), & RG_j \text{ is cost attribute} \end{cases} \quad (8)$$

Step 3. Administrate the DVNN negative ideal decision solution (DVNNNIDS):

$$DVNNNIDS = \begin{pmatrix} DVNNNIDS_1, DVNNNIDS_2, \\ \cdots, DVNNNIDS_n \end{pmatrix} \quad (9)$$

$$DVNNNIDS_j = \left(RT_j^{N-}, RIT_j^{N-}, RIF_j^{N-}, RF_j^{N-}\right) \quad (10)$$

$$DVNNSV\left(DVNNNIDS_j\right) = \min_i DVNNSV\left(RT_{ij}^{N}, RIT_{ij}^{N}, RIF_{ij}^{N}, RF_{ij}^{N}\right) \quad (11)$$

Step 4. Administrate the attribute weight through CRITIC.

The CRITIC [37] is administrated to obtain the objective weight.

1) Administrate the DVNNED and DVNNLD of RA_i from DVNNNIDS.

$$DVNNED\left(NRM_{ij}, DVNNNIDS_j\right) = \sqrt{\frac{1}{4}\left(\begin{array}{l}\left|RT_{ij}^{N} - RT_{j}^{N-}\right|^2 + \left|RIT_{ij}^{N} - RIT_{j}^{N-}\right|^2 \\ + \left|RIF_{ij}^{N} - RIF_{j}^{N-}\right|^2 + \left|RF_{ij}^{N} - RF_{j}^{N-}\right|^2\end{array}\right)} \quad (12)$$

$$DVNNLD\left(NRM_{ij}, DVNNNIDS_j\right) = \frac{1}{4}\left(\begin{array}{l} RT_{ij}^{N}\log\frac{2RT_{ij}^{N}}{RT_{ij}^{N}+RT_{j}^{N-}} + RT_{j}^{N-}\log\frac{2RT_{j}^{N-}}{RT_{ij}^{N}+RT_{j}^{N-}} \\ + RIT_{ij}^{N}\log\frac{2RIT_{ij}^{N}}{RIT_{ij}^{N}+RIT_{j}^{N-}} + RIT_{j}^{N-}\log\frac{2RIT_{j}^{N-}}{RIT_{ij}^{N}+RIT_{j}^{N-}} \\ + RIF_{ij}^{N}\log\frac{2RIF_{ij}^{N}}{RIF_{ij}^{N}+RIF_{j}^{N-}} + RIF_{j}^{N-}\log\frac{2RIF_{j}^{N-}}{RIF_{ij}^{N}+RIF_{j}^{N-}} \\ + RF_{ij}^{N}\log\frac{2RF_{ij}^{N}}{RF_{ij}^{N}+RF_{j}^{N-}} + RF_{j}^{N-}\log\frac{2RF_{j}^{N-}}{RF_{ij}^{N}+RF_{j}^{N-}}\end{array}\right) \quad (13)$$

$$NRMRA_j = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{\left(\begin{array}{l}DVNNED\left(NRM_{ij}, DVNNNIS_j\right) \\ + DVNNLD\left(NRM_{ij}, DVNNNIDS_j\right)\end{array}\right)}{2}\right), j = 1, 2, \cdots, n. \quad (14)$$

$$NHHOA_{ij} = \frac{\left(\begin{array}{l}DVNNED\left(NRM_{ij}, DVNNNIS_j\right) \\ + DVNNLD\left(NRM_{ij}, DVNNNIDS_j\right)\end{array}\right)}{2} \quad (15)$$

2) Administrate the DVNN correlation decision coefficient (DVNNCDC):

$$DVNNCDC_{jk} = \frac{\sum_{i=1}^{m}\left(NRMRA_{ij} - NRMRA_j\right)\left(NRMRA_{ik} - NRMRA_k\right)}{\sqrt{\sum_{i=1}^{m}\left(NRMRA_{ij} - NRMRA_j\right)^2\sum_{i=1}^{m}\left(NRMRA_{ik} - NRMRA_k\right)^2}} \quad (16)$$

3) Administrate the DVNN standard deviation ($DVNNSD_j$):

$$DVNNSD_j = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}\left(NRMRA_{ij} - NRMRA_j\right)^2}, \quad (17)$$

(4) Administrate the weight numbers:

$$rw_j = \frac{DVNNSD_j\left(1 - DVNNCDC_{jk}\right)}{\sum_{k=1}^{n} DVNNSD_j\left(1 - DVNNCDC_{jk}\right)} \quad (18)$$

Step 5. Administrate the DVNN weighted averaging information (DVNNWAI).

$$DVNNWAI_i = \left(\sum_{j=1}^{n} \left(rw_j \times \begin{pmatrix} DVNNED\left(NHH_{ij}, DVNNNIS_j\right) \\ +DVNNLD\left(NRM_{ij}, DVNNNIDS_j\right) \end{pmatrix} \right) \Big/ 2 \right). \tag{19}$$

Step 6. Administrate the DVNN weighted geometric information (DVNNWGI).

$$DVNNWGI_i = \prod_{j=1}^{n} \left(\left(\begin{pmatrix} DVNNED\left(NHH_{ij}, DVNNNIS_j\right) \\ +DVNNLD\left(NRM_{ij}, DVNNNIDS_j\right) \end{pmatrix} \right) \Big/ 2 \right)^{rw_j}. \tag{20}$$

Step 7. Administrate the three sorting strategies for relative importance.

$$DVNNR_{ia} = \frac{DVNNWGI_i + DVNNWAI_i}{\sum_{i=1}^{m} \left(DVNNWGI_i + DVNNWAI_i \right)}. \tag{21}$$

$$DVNNR_{ib} = \frac{DVNNWAI_i}{\min\limits_i DVNNWAI_i} + \frac{DVNNWGI_i}{\min\limits_i DVNNWGI_i}. \tag{22}$$

$$DVNNR_{ic} = \frac{DVNNWAI_i + (1-\lambda) DVNNWGI_i}{\lambda \max\limits_i DVNNWAI_i + (1-\lambda) \max\limits_i DVNNWGI_i}, 0 \le \lambda \le 1. \tag{23}$$

where $DVNNR_{ia}$ is arithmetic average of $DVNNWAI_i, DVNNWGI$, $DVNNR_{ib}$ is relative scores of $DVNNWAI_i, DVNNWGI$, and $DVNNR_{ic}$ is balanced compromise of $DVNNWAI_i, DVNNWGI$.

Step 8. Administrate the sorted values $DVNNR_i$.

$$DVNNR_i = \begin{pmatrix} \sqrt[3]{DVNNR_{ia} DVNNR_{ib} DVNNR_{ic}} \\ +\dfrac{DVNNR_{ia} + DVNNR_{ib} + DVNNR_{ic}}{3} \end{pmatrix} \tag{24}$$

Step 9. Administrate the alternatives with $DVNNR_i \left(i = 1, 2, \cdots, m\right)$, and higher $DVNNR_i$ is better alternative.

IV. Example Study and Comparative Analysis

A. Example Study for Management Performance Evaluation of Teaching Services in Sports Colleges and Universities

The current teaching management in universities is becoming increasingly complex, and some teaching management personnel have deficiencies in concepts, professional qualities, knowledge, and mentality, which leads to a certain degree of insufficient human resources in the actual operation process of teaching management [38-40]. Collaboration and teamwork between teaching management departments are beneficial for members to overcome their own knowledge limitations, which can be achieved through rotation of teaching management positions or team work based on the time period of teaching management work. The reasonable flow of teaching management personnel within the department is an effective way for personal knowledge to be shared within the organization, which is conducive to knowledge transfer among members [41, 42]. The teaching management work of schools is mainly focused on supporting teaching work, and the nature of services is very obvious. In the practice of teaching management, teaching managers should establish the ideological consciousness of serving teaching work, serving grassroots, teaching, teachers, and students, and closely focus on providing various services around teaching work, striving to improve the level of management services [43-46]. To some extent, the process of service is the process of knowledge transfer. Teaching managers can openly express their opinions, maintain efficient communication with other managers, teachers, and students, promote the integration of various information in business relationships, and improve the behavioral paradigm of teaching managers. Continuously learning in the management process, emphasizing the learning process, summarizing the problems or advantages discovered in practice, and internalizing the correct experiences of these learning outcomes into organizational memory [47-49]. By establishing a long-term learning mechanism, a common communication channel is created for explicit and implicit knowledge within the organization, forming a shared learning system and accelerating the flow of knowledge within the organization [17, 50, 51]. We will adhere to the methods of development and reform to solve the problems in progress,

vigorously promote innovation in the teaching management system and mechanism, further improve the level of teaching management, enhance the awareness and quality of teaching management services, and achieve the goal of cultivating high-quality composite applied talents with innovative spirit and practical ability [52-55]. Education, scientific research, and training will develop comprehensively, coordinated, and sustainably. People use defensive reasoning for thinking and action in their daily lives and work because they have developed this habit for a long time in such an environment. Habitual defense is an important protective factor that prevents people from bravely facing and correcting mistakes, and hinders organizations from continuously learning and developing. In the process of teaching management reform in schools, we often encounter such problems. In the face of either new and old systems or innovation, we both face the dilemma of autonomy and control [56-58]. So, self-protection emerged, with a vague and unclear attitude towards new things, fearing that change would harm one's own interests. Establishing a creative learning environment in teaching management practice, improving the reflective learning ability of management personnel, and promoting the continuous development of teaching management [59]. The management performance evaluation of teaching services in sports colleges and universities is MADM. There are five possible sports colleges and universities $RA_i\ (i=1,2,3,4,5)$ are assessed in light with four attributes: ①RG_1 is the student feedback results; ②RG_2 is the blended teaching management costs; ③RG_3 is the blended teaching attitude; ④RG_4 is the invited peer review recognition. RG_2 is the cost. Then, the DVNN-CoCoSo model is administrated for management performance evaluation of teaching services in sports colleges and universities.

Step 1. Put forward the DVNN-matrix $RM=\left(UR_{ij}\right)_{5\times4}$ as in Table I.

TABLE I. DVNN INFORMATION

	RG_1	RG_2
RA_1	(0.35,0.64, 0.13, 0.46)	(0.42, 0.36, 0.45, 0.38)
RA_2	(0.53, 0.42,0.35, 0.54)	(0.63, 0.56, 0.37, 0.42)
RA_3	(0.46, 0.35, 0.49,0.43)	(0.29, 0.35, 0.46, 0.24)
RA_4	(0.43, 0.34, 0.37, 0.42)	(0.34, 0.53, 0.42, 0.46)
RA_5	(0.35, 0.46,0.39, 0.37)	(0.37, 0.59, 0.36, 0.29)
	RG_3	RG_4
RA_1	(0.49, 0.26, 0.58, 0.45)	(0.36, 0.38, 0.23, 0.43)
RA_2	(0.42, 0.31,0.52, 0.43)	(0.42, 0.57,0.16, 0.45)
RA_3	(0.25, 0.39,0.37, 0.46)	(0.49, 0.54,0.32, 0.41)
RA_4	(0.32, 0.46, 0.45, 0.39)	(0.35, 0.51,0.39, 0.42)
RA_5	(0.34, 0.48,0.42, 0.37)	(0.37, 0.64,0.15, 0.48)

Step 2. Normalize the DVNN matrix $RM=\left(UR_{ij}\right)_{5\times4}$ to $NRM=\left(NRM_{ij}\right)_{5\times4}$ (see Table II).

TABLE II. THE NORMALIZED DVNNS

	RG_1	RG_2
RA_1	(0.25, 0.39,0.37, 0.46)	(0.41, 0.32,0.54, 0.49)
RA_2	(0.32, 0.46, 0.45, 0.39)	(0.42, 0.39,0.51, 0.35)
RA_3	(0.34, 0.48,0.42, 0.37)	(0.48, 0.15,0.64, 0.37)
RA_4	(0.49, 0.26, 0.58, 0.45)	(0.43, 0.23, 0.38, 0.36)
RA_5	(0.42, 0.31,0.52, 0.43)	(0.45, 0.16,0.57, 0.42)
	RG_3	RG_4
RA_1	(0.43, 0.34, 0.37, 0.42)	(0.34, 0.53, 0.42, 0.46)
RA_2	(0.35, 0.46,0.39, 0.37)	(0.37, 0.59, 0.36, 0.29)
RA_3	(0.35,0.64, 0.13, 0.46)	(0.42, 0.36, 0.45, 0.38)
RA_4	(0.53, 0.42,0.35, 0.54)	(0.63, 0.56, 0.37, 0.42)
RA_5	(0.46, 0.35, 0.49,0.43)	(0.29, 0.35, 0.46, 0.24)

Step 3. Administrate the weight numbers (Table III).

TABLE III. THE WEIGHT NUMBERS

	RG_1	RG_2	RG_3	RG_4
weight	0.2207	0.3028	0.2504	0.2261

Step 4. Put forward the DVNNWAI (Table IV).

TABLE IV. THE DVNNWAI

	RA_1	RA_2	RA_3	RA_4	RA_5
DVNNWAI	0.8145	0.5597	0.6728	0.6984	0.4652

Step 5. Administrate the DVNNWGI (Table V).

TABLE V. THE DVNNWGI

	RA_1	RA_2	RA_3	RA_4	RA_5
DVNNWGI	0.7378	0.4613	0.5736	0.5879	0.3657

Step 6. Administrate the $DVNNR_{ia}, DVNNR_{ib}, DVNNR_{ic}$ (see Table VI).

TABLE VI. THREE DECISION STRATEGIES

	$DVNNR_{ia}$	$DVNNR_{ib}$	$DVNNR_{ic}$
RA_1	0.2615	3.7684	1.0000
RA_2	0.1720	2.4646	0.6577
RA_3	0.2099	3.0148	0.8029
RA_4	0.2167	3.1089	0.8286
RA_5	0.1400	2.0000	0.5353

Step 7. Administrate the DVNNR (see Table VII).

TABLE VII. THE DVNNR

	RA_1	RA_2	RA_3	RA_4	RA_5
DVNNR	2.6717	1.7513	2.1406	2.2081	1.4229

Step 8. According to $DVNNR_i\left(i=1,2,3,4,5\right)$, the order is $RA_1 > RA_4 > RA_3 > RA_2 > RA_5$ and the optimal sports college and university is RA_1.

B. Comparative Analysis

The DVNN-CoCoSo technique is compared with generalized double-valued neutrosophic weighted distance [36] and weighted Dice similarity measures $WD_{\mathrm{DVNS}_1}\left(RA_i, DVNNPIS\right)$, $WD_{\mathrm{DVNS}_2}\left(RA_i, DVNNPIS\right)$ and weighted generalized Dice similarity measures $WGD_{\mathrm{DVNS}_1}\left(RA_i, DVNNPIS\right)$, $WGD_{\mathrm{DVNS}_2}\left(RA_i, DVNNPIS\right)$ [60], DVNN-TODIM-VIKOR technique [61] and DVNN-ExpTODIM-GRA technique [62]. The comparative results are administrated in Table VIII.

TABLE VIII. ORDER FOR DIFFERENT TECHNIQUES

	Order
DVNN weighted Hamming distance[36]	$RA_1 > RA_4 > RA_3 > RA_2 > RA_5$
DVNN weighted Euclidean distance[36]	$RA_1 > RA_4 > RA_3 > RA_2 > RA_5$
$WD_{\mathrm{DVNS}_1}\left(RA_i, DVNNPIS\right)$ [60]	$RA_1 > RA_4 > RA_3 > RA_2 > RA_5$
$WD_{\mathrm{DVNS}_2}\left(RA_i, DVNNPIS\right)$ [60]	$RA_1 > RA_4 > RA_3 > RA_2 > RA_5$
$WGD_{\mathrm{DVNS}_1}\left(RA_i, DVNNPIS\right)$ [60]	$RA_1 > RA_4 > RA_3 > RA_2 > RA_5$
$WGD_{\mathrm{DVNS}_2}\left(RA_i, DVNNPIS\right)$ [60]	$RA_1 > RA_4 > RA_3 > RA_2 > RA_5$
DVNN-TODIM-VIKOR technique [61]	$RA_1 > RA_4 > RA_2 > RA_3 > RA_5$
DVNN-ExpTODIM-GRA technique [62]	$RA_1 > RA_4 > RA_2 > RA_3 > RA_5$
DVNN-CoCoSo technique	$RA_1 > RA_4 > RA_3 > RA_2 > RA_5$

From the above comparative analysis, the order of generalized double-valued neutrosophic weighted distance [36] and weighted Dice similarity measures $WD_{\mathrm{DVNS}_1}\left(RA_i, DVNNPIS\right)$, $WD_{\mathrm{DVNS}_2}\left(RA_i, DVNNPIS\right)$ and weighted generalized Dice similarity measures $WGD_{\mathrm{DVNS}_1}\left(RA_i, DVNNPIS\right)$, $WGD_{\mathrm{DVNS}_2}\left(RA_i, DVNNPIS\right)$ [60] is same to order of DVNN-CoCoSo technique; while order of DVNN-TODIM-VIKOR technique [61] and DVNN-ExpTODIM-GRA technique [62] is slightly different from order of DVNN-CoCoSo technique, however, several techniques have same optimal sports college and university and worst sports college and university. This verifies the effectiveness of DVNN-CoCoSo technique. Thus, the main advantages of DVNN-CoCoSo are managed: (1) DVNN-CoCoSo technique not only manages the uncertainty for MADM, but also manages three fused strategies. (2) DVNN-CoCoSo manages different behavior of CoCoSo and entropy as MADM when they are combined.

V. CONCLUSION

The service performance of the teaching management department is an important basis for measuring and evaluating the quality and operational efficiency of teaching management in higher education institutions, and is a comprehensive reflection of the school's management capabilities. Sports colleges and universities are the highest level of education in China's higher education system to cultivate sports professionals, shouldering the arduous task of cultivating sports talents with innovative spirit and practical ability, and contributing to the country's sports and education undertakings. The service performance construction of the teaching management department is achieved by adopting a scientific management approach, stimulating the enthusiasm, initiative, and creativity of teaching staff, correcting service attitudes, enhancing the consciousness of actively serving teaching work, thereby improving work efficiency and ability, and better implementing the school's policies and guidelines. The management performance evaluation of teaching services in sports colleges and universities is regarded as MADM. Consequently, the DVNN-CoCoSo technique is administrated to put forward MADM for management performance evaluation of teaching services in sports colleges and universities. The major contribution of this paper is administrated: 1) DVNN-CoCoSo technique is administrated for MADM in light with DVNNED and DVNNLD; 2) the objective weights are considered through entropy technique; 3) numerical example for management performance evaluation of teaching services in sports colleges and universities and some comparative analysis are administrated to verify the DVNN-CoCoSo technique.

There may be some possible limitations for management performance evaluation of teaching services in sports colleges and universities, which could be conducted through our future research: 1) It is a worthwhile research work to manage consensus [63-66] for management performance evaluation of teaching services in sports colleges and universities under DVNSs. 2) It is also worthwhile research to manage regret theory for management performance evaluation of teaching services in sports colleges and universities under DVNSs [67-70].

REFERENCES

[1] W. Jing, A study on the evaluation model of teaching quality of accounting majors under the perspective of sports psychology, Revista De Psicologia Del Deporte, 31 (2022) 49-61.

[2] T.T. Qiu, D.L. Zhou, W. Li, Fitness culture and green space equity: Accessibility evaluation of shanghai communities, Frontiers in Environmental Science, 10 (2022) 9.

[3] H.M. Wang, M. Chen, Application of the flipped classroom mode under few-shot learning in the teaching of health physical education in colleges and universities, Computational Intelligence and Neuroscience, 2022 (2022) 10.

[4] J.Q. Wang, D.M. Shi, S.B. Tsai, An empirical study on the construction path of sports public service platform under the background of "internet plus " by fuzzy computing, Mathematical Problems in Engineering, 2022 (2022) 10.

[5] S. Wang, Influence of management efficiency of sports equipment in colleges and universities based on the intelligent optimization method, Scientific Programming, 2022 (2022) 8.

[6] Z. Wang, Y.H. Liu, S. Zhang, Multisensor data fusion system for wushu sanda teaching in higher education institutions, Wireless Communications & Mobile Computing, 2022 (2022) 10.

[7] D. Wu, Z.L. Song, H. Guo, Artificial intelligence algorithms in ice and snow tourism promotion from digital technology, Wireless Communications & Mobile Computing, 2022 (2022) 9.

[8] L. Zhang, Construction of evaluation indicator system of college physical education teaching environment based on analytic hierarchy process, Computational Intelligence and Neuroscience, 2022 (2022) 10.

[9] Q.X. Zhang, Construction of leisure physical education teaching model based on multisensor fusion, Mobile Information Systems, 2022 (2022) 10.

[10] F. Guo, Construction of intelligent supervision platform for college students' physical health for intelligent medical service decision-making, Frontiers in Physics, 11 (2023) 10.

[11] G. Jones, E. Macaninch, D.D. Mellor, A. Spiro, K. Martyn, T. Butler, A. Johnson, J.B. Moore, Putting nutrition education on the table: Development of a curriculum to meet future doctors' needs, British Journal of Nutrition, 129 (2023) 1000-1008.

[12] J. Li, S.Y. Tian, S. Charoenwattana, Smart iot-based visual target enabled track and field training using image recognition, Soft Computing, 27 (2023) 12571-12585.

[13] S.C. Li, C.F. Tang, C. Guo, T. Bu, Psychometric properties of the sport anxiety scale-2 for chinese adolescent athletes taking the national sports college entrance examination, Frontiers in Pediatrics, 11 (2023) 11.

[14] X.L. Li, Optimization of the college basketball teaching mode based on the applied explainable association rule algorithm and cluster analysis in mobile computing environments, Applied Artificial Intelligence, 37 (2023) 22.

[15] M.E. Marshall-Gibson, M.G. Durham, K.A. Seabaugh, V.J. Moorman, D.J. Ferris, Survey of equine veterinarians regarding primary equine back pain in the united states, Frontiers in Veterinary Science, 10 (2023) 10.

[16] S. Riciputi, M.H. McDonough, F.J. Snyder, M.L. McDavid, Staff support promotes engagement in a physical activity-based positive youth development program for youth from low-income families, Sport Exercise and Performance Psychology, 9 (2020) 45-57.

[17] T. Sato, E. Tsuda, D. Ellison, S.R. Hodge, Japanese elementary teachers' professional development experiences in physical education lesson studies, Physical Education and Sport Pedagogy, 25 (2020) 137-153.

[18] M.M. Alhumaid, S. Khoo, T. Bastos, The effect of an adapted physical activity intervention program on pre-service physical education teachers' self-efficacy towards inclusion in saudi arabia, Sustainability, 13 (2021) 13.

[19] S. Park, H.W. Lee, Emphasizing effort vs talent in personal trainers' performance: Consumption response of personal fitness training customers, International Journal of Sports Marketing & Sponsorship, 24 (2023) 359-374.

[20] M. Simon, K. Boyd, Cracks in the narrative: Black and latinx pre-service pe teachers in predominantly white pete programs, Physical Education and Sport Pedagogy, 28 (2023) 259-275.

[21] T.T. Wang, Effect of using 5g and cloud computing environment for independent college english vocabulary learning, Journal of Cloud Computing-Advances Systems and Applications, 12 (2023) 12.

[22] S.G. Kennedy, J.J. Smith, P.A. Estabrooks, N. Nathan, M. Noetel, P.J. Morgan, J. Salmon, G.C. Dos Santos, D.R. Lubans, Evaluating the reach, effectiveness, adoption, implementation and maintenance of the resistance training for teens program, International Journal of Behavioral Nutrition and Physical Activity, 18 (2021) 18.

[23] A.C. Odole, O.T. Agbomeji, O.K.K. Onyeso, J.O. Ojo, N.A. Odunaiya, Perspectives of nigerian athletes about physiotherapy services in sports injury management: Implications for rehabilitation, Journal of Sport Rehabilitation, 30 (2021) 876-883.

[24] A. Soini, A. Watt, A. Sääkslahti, Finnish pre-service teachers' perceptions of perceived competence in early childhood physical education, International Journal of Environmental Research and Public Health, 18 (2021) 14.

[25] A.L. Yang, Y.S. Wang, L. Yao, X.D. Hamilton, Revalidation of preservice physical education teacher's teacher identity scale in chinese physical education student teachers, Frontiers in Psychology, 13 (2023) 12.

[26] Y.C. Zhang, W.T. Duan, L.E. Villanueva, S. Chen, Transforming sports training through the integration of internet technology and artificial intelligence, Soft Computing, 27 (2023) 15409-15423.

[27] R. Burgueno, A. Abós, J. Sevil-Serrano, L. Haerens, K. De Cocker, L. García-González, A circumplex approach to (de)motivating styles in physical education: Situations-in-school-physical education questionnaire in spanish students, pre-service, and in-service teachers, Measurement in Physical Education and Exercise Science, 28 (2024) 86-108.

[28] L. Hasson, S. Keville, J. Gallagher, D. Onagbesan, A.K. Ludlow, Inclusivity in education for autism spectrum disorders: Experiences of support from the perspective of parent/carers, school teaching staff and young people on the autism spectrum, International Journal of Developmental Disabilities, 70 (2024) 201-212.

[29] Q.F. Yu, B.S. Liu, J.L. Zang, S.J. Wang, The reform of supply of public health services leading the training of sports professionals in local colleges and universities in the background of healthy china, Revista Brasileira De Medicina Do Esporte, 27 (2021) 101-104.

[30] K. Zhang, H.Y. Wang, J.Y. Wang, C.L. Li, J.Y. Zhang, Nature-based experiential learning as a framework for preparing responsible tourism practitioners, Journal of Hospitality Leisure Sport & Tourism Education, 29 (2021) 9.

[31] P. Liu, D. Shi, B. Zang, X. Liu, Students health physique information sharing in publicly collaborative services over edge-cloud networks, Journal of Cloud Computing-Advances Systems and Applications, 13 (2024) 11.

[32] J. Wang, Y.L. Yang, H.Y. Liu, L.Z. Jiang, Enhancing the college and university physical education teaching and learning experience using virtual reality and particle swarm optimization, Soft Computing, 28 (2024) 1653-1667.

[33] J.H. Xu, D.L. Qi, S. Liu, Intelligent sports teaching tracking system based on multimedia data analysis and artificial intelligence, Tehnicki Vjesnik-Technical Gazette, 31 (2024) 951-958.

[34] M. Yazdani, P. Zarate, E.K. Zavadskas, Z. Turskis, A combined compromise solution (cocoso) method for multi-criteria decision-making problems., Management Decision, 57 (2018) 2501-2519.

[35] C.E. Shannon, A mathematical theory of communication, Bell System Technical Journal, 27 (1948) 379-423.

[36] I. Kandasamy, Double-valued neutrosophic sets, their minimum spanning trees, and clustering algorithm, Journal of Intelligent Systems, 27 (2018) 163-182.

[37] D. Diakoulaki, G. Mavrotas, L. Papayannakis, Determining objective weights in multiple criteria problems the critic method, Elsevier Scienc, 22 (1995) 767-770.

[38] H.Q. Qiu, Q.H. Li, C.X. Li, How technology facilitates tourism education in covid-19:Case study of nankai university, Journal of Hospitality Leisure Sport & Tourism Education, 29 (2021) 8.

[39] D. Shan, The construction of a professional model of physical education teachers in colleges and universities from the perspective of public health service, Revista Brasileira De Medicina Do Esporte, 27 (2021) 59-61.

[40] X.Y. Xie, P. Ward, D. Oh, Y.L. Li, O. Atkinson, K. Cho, M. Kim, Preservice physical education teacher's development of adaptive competence, Journal of Teaching in Physical Education, 40 (2021) 538-546.

[41] W.M. Li, X.R. Fan, Construction of network multimedia teaching platform system of college sports, Mathematical Problems in Engineering, 2021 (2021) 11.

[42] L.S. Liu, Optimizing kenmi manipulation courses of high school sports based on cdio model under the background of cloud computing, Scientific Programming, 2021 (2021) 5.

[43] J.L. Walton-Fisette, T.A. Walton-Fisette, L.F. Chase, Captured on film: A critical examination of representations of physical education at the movies, Physical Education and Sport Pedagogy, 22 (2017) 536-547.

[44] J.H. Baek, E. Jones, S. Bulger, A. Taliaferro, Physical education teacher perceptions of technology-related learning experiences: A qualitative investigation, Journal of Teaching in Physical Education, 37 (2018) 175-185.

[45] O. Furtado, L.J. Lieberman, G.L. Gutierrez, The outcomes of running a sport camp for children and youth with visual impairments on faculty members' teaching, research, and service activities: A case study, Sport in Society, 21 (2018) 76-90.

[46] J. LaGrandeur, M. Moros, J. Dobrick, R. Rahimian, A. Siyahian, E. Tomlinson, P. Gordon, Totshots: An innovative pediatric free clinic providing high patient satisfaction to the underserved, Family Medicine, 50 (2018) 779-781.

[47] S. Harvey, C. Cushion, P. Sammon, Dilemmas faced by pre-service teachers when learning about and implementing a game-centred approach, European Physical Education Review, 21 (2015) 238-256.

[48] E. Enright, L. Coll, D. Ní Chróinín, M. Fitzpatrick, Student voice as risky praxis: Democratising physical education teacher education, Physical Education and Sport Pedagogy, 22 (2017) 459-472.

[49] E.M. Jones, J.H. Baek, J.D. Wyant, Exploring pre-service physical education teacher technology use during student teaching, Journal of Teaching in Physical Education, 36 (2017) 173-184.

[50] B. Zhang, D.X. Chen, Resource scheduling of green communication network for large sports events based on edge computing, Computer Communications, 159 (2020) 299-309.

[51] A. Bramley, A. Forsyth, L. McKenna, Design, implementation and evaluation of novel work-based clinical assessment tool: An e-portfolio with embedded entrustable professional activities, Nurse Education Today, 107 (2021) 7.

[52] R. Marttinen, D.N. Daum, D. Banville, R.N. Fredrick, Pre-service teachers learning through service-learning in a low ses school, Physical Education and Sport Pedagogy, 25 (2020) 1-15.

[53] R. Cohen, J.D. Goodway, R. Lidor, The effectiveness of aligned developmental feedback on the overhand throw in third-grade students, Physical Education and Sport Pedagogy, 17 (2012) 525-541.

[54] S. Zach, I. Harari, N. Harari, Changes in teaching efficacy of pre-service teachers in physical education, Physical Education and Sport Pedagogy, 17 (2012) 447-462.

[55] D. Özer, S. Nalbant, E. Aglamis, F. Baran, P.K. Samut, A. Aktop, Y. Hutzler, Physical education teachers' attitudes towards children with intellectual disability: The impact of time in service, gender, and previous acquaintance, Journal of Intellectual Disability Research, 57 (2013) 1001-1013.

[56] C. Luguetti, D. Kirk, K.L. Oliver, Towards a pedagogy of love: Exploring pre-service teachers' and youth's experiences of an activist sport pedagogical model, Physical Education and Sport Pedagogy, 24 (2019) 629-646.

[57] Y.D. Yuan, Z. Li, H.W. Liang, Selection and spatial layout of urban sports facilities guided by residents' demand, Open House International, 44 (2019) 9-12.

[58] L. Du, Q.F. Yu, The development path of coastal sports tourism industry from the perspective of ecological protection, Journal of Coastal Research, (2020) 120-123.

[59] J. Qi, L.J. Wang, Social interaction between students with and without disabilities in general physical education: A chinese perspective, Physical Education and Sport Pedagogy, 23 (2018) 575-591.

[60] Q. Khan, P. Liu, T. Mahmood, Some generalized dice measures for double-valued neutrosophic sets and their applications, Mathematics, 6 (2018) 121.

[61] K. Du, Y. Du, Research on management performance evaluation of intangible assets operation and management in sports events with double-valued neutrosophic sets, Journal of Intelligent & Fuzzy Systems, 45 (2023) 2813-2822.

[62] M. Gong, Fuzzy multiple attribute decision making method for multimedia teaching effectiveness comprehensive evaluation in college english with double-valued neutrosophic numbers, Journal of Intelligent & Fuzzy Systems, 45 (2023) 5697-5707.

[63] P. Wu, F.G. Li, J. Zhao, L.G. Zhou, L. Martfnez, Consensus reaching process with multiobjective optimization for large-scale group decision making with cooperative game, Ieee Transactions on Fuzzy Systems, 31 (2023) 293-306.

[64] X.X. Xu, Z.W. Gong, E. Herrera-Viedma, G. Kou, F.J. Cabrerizo, Consensus reaching in group decision making with linear uncertain preferences and asymmetric costs, Ieee Transactions on Systems Man Cybernetics-Systems, 53 (2023) 2887-2899.

[65] H.M. Zhang, Y.Y. Dai, Consensus improvement model in group decision making with hesitant fuzzy linguistic term sets or hesitant fuzzy linguistic preference relations, Computers & Industrial Engineering, 178 (2023) 14.

[66] H. Nakase, M. Esaki, F. Hirai, T. Kobayashi, K. Matsuoka, M. Matsuura, M. Naganuma, M. Saruta, K. Tsuchiya, M. Uchino, K. Watanabe, T. Hisamatsu, T.C. Grp, Treatment escalation and de-escalation decisions in crohn's disease: Delphi consensus recommendations from japan, 2021, Journal of Gastroenterology, 58 (2023) 313-345.

[67] X.L. Tian, Z.S. Xu, J. Gu, F. Herrera, A consensus process based on regret theory with probabilistic linguistic term sets and its application in venture capital, Information Sciences, 562 (2021) 347-369.

[68] H.P. Ren, Y.X. Gao, T.H. Yang, A novel regret theory-based decision-making method combined with the intuitionistic fuzzy canberra distance, Discrete Dynamics in Nature and Society, 2020 (2020) 9.

[69] X. Jia, X.F. Wang, Y.F. Zhu, L. Zhou, H. Zhou, A two-sided matching decision-making approach based on regret theory under intuitionistic fuzzy environment, Journal of Intelligent & Fuzzy Systems, 40 (2021) 11491-11508.

[70] Y. Lin, Y.M. Wang, S.Q. Chen, Hesitant fuzzy multiattribute matching decision making based on regret theory with uncertain weights, International Journal of Fuzzy Systems, 19 (2017) 955-966.

Enhanced Methodology for Production-Education Integration and Quality Evaluation of Rural Vocational Education Under Rural Revitalization with 2-Tuple Linguistic Neutrosophic Numbers

Xingli Wang*
School of Marxism, Shenzhen Polytechnic University, Shenzhen 518055, Guangdong, China

***Abstract*—Rural vocational education (RVE) plays a crucial role in nurturing practical talents for the development of rural economy and society in the new era, as well as cultivating future generations of agricultural successors. Quality evaluation serves as an essential means of ensuring educational excellence. It acts as a key element for supervision, assurance, and enhancement of educational quality. For the production-education integration (PEI) in RVE, establishing a quality evaluation system that aligns with national and rural conditions, caters to the needs of modern agricultural industry development, and reflects the characteristics of RVE is crucial. Such a system plays a vital role in leading and promoting the deep integration of industry and education in rural vocational colleges. The PEI quality evaluation in RVE under rural revitalization involves MAGDM. Currently, Exponential TODIM (ExpTODIM) approach and grey relational analysis (GRA) approach has been utilized to address MAGDM challenges. To handle uncertain information in PEI quality evaluation in RVE under rural revitalization, 2-tuple linguistic neutrosophic sets (2TLNSs) are conducted as a valuable tool. This paper introduces the implementation of the 2-tuple linguistic neutrosophic number Exponential TODIM-GRA (2TLNN-ExpTODIM-GRA) approach to effectively manage MAGDM problems using 2TLNSs. Additionally, a numerical study is conducted to validate the application of this approach for PEI quality evaluation in RVE under rural revitalization.**

Keywords*—Multiple-attribute group decision-making (MAGDM); 2TLNSs; ExpTODIM approach; GRA approach; PEI quality evaluation

I. Introduction

The adaptability of RVE to the needs of rural economic and social development in the new era is a prerequisite for RVE to fully play its inherent role and effectively serve rural economic and social development [1-3]. "Adaptability" means changing with the external environment. Under the rural revitalization strategy, development of the rural economy and society in the new era, agricultural modernization, and new urbanization construction, external development environment of RVE has undergone tremendous changes [4, 5]. This poses new tasks and requirements for RVE, requiring it to adapt to external environment changes and times demands, and timely adjust and optimize its positioning, development ideas, program offerings, and talent development goals to meet the changing external environment and the requirements of the times[6, 7]. Therefore, to assess and evaluate the adaptability of RVE to the needs of rural economic development and social development for a new era, it is necessary to examine the positioning of agricultural colleges and the positioning of talent development goals from the perspective of external demand. It is important to monitor the compatibility between the development plans of agricultural colleges and rural revitalization strategy, rural economic and social development, agricultural modernization, urbanization construction, and local agricultural development and rural development plans [8, 9]. It is necessary to observe whether the development plans are synchronized with the new era's rural economic and social development plans [10, 11]. It is crucial to observe the alignment between the professional development plans of agricultural colleges and local industries, especially the relationship between the school's advantages, distinctive programs, and leading majors, and the layout and structure of local modern agricultural industries [12, 13]. The construction of programs should closely follow the dynamic adjustment mechanism of local agricultural industry transformation and upgrading. It is essential to observe the alignment between the talent development goals of colleges and the needs of various types of rural talents and their qualifications and structure required for rural revitalization [14, 15]. The talent development goals and specifications of each major should reflect the need for compound and innovative rural talents with systematic knowledge structures and comprehensive abilities to adapt to new industries development, new formats, and new modes for rural areas [16-18].

Decision-making is a deliberate and selective process employed by humans to accomplish specific objectives [19-22]. Essentially, it involves utilizing available decision information, employing a particular method to evaluate various alternative solutions, and selecting the best solution that aligns with the decision-maker's expectations [23-27]. In our daily lives, decision-making is intertwined with various aspects such as food, clothing, housing, and transportation, where we constantly make choices [28-35]. The PEI quality evaluation of RVE under rural revitalization involves MAGDM. To address MAGDM challenges, recent approaches have utilized the ExpTODIM approach [36, 37] and GRA approach [38] approach [37]. Additionally, 2TLNSs [39] have been administrated to represent uncertain information during the PEI quality evaluation of RVE under rural revitalization. This paper introduces the ExpTODIM-GRA approach designed

specifically to manage MAGDM problems using 2TLNSs. Firstly, the basic concepts of 2TLNSs are reviewed. Then, ExpTODIM-GRA approach is administrated to address MAGDM under 2TLNSs. Finally, numerical study is administrated to validate the effectiveness of 2TLNN-ExpTODIM-GRA in the PEI quality evaluation of RVE under rural revitalization. In summary, the motivation behind this study is administrated: (1) the utilization of Entropy [39] to obtain attribute weights with 2TLNSs; (2) the implementation of the 2TLNN-ExpTODIM-GRA approach to manage the MAGDM with 2TLNSs; (3) the construction of numerical study to demonstrate the application of the administrated approach in the PEI quality evaluation of RVE under rural revitalization; and (4) the provision of comparative analyses with existing approaches.

The organization of this work is conducted: Section II introduces 2TLNSs. Section III describes the implementation of 2TLNN-ExpTODIM-GRA approach under the framework of 2TLNSs, incorporating the entropy approach. In Section IV, numerical example is presented to showcase the PEI quality evaluation in RVE under rural revitalization, accompanied by a comparative analysis. Finally, Section V offers concluding remarks to wrap up the study.

II. PRELIMINARIES

Wang et al. [39] administrated the 2TLNSs.

Definition 1[28-29]**.** Let $ds_1, ds_2, \ldots, ds_\phi$ be linguistic information and ds is administrated:

$$ds = \begin{Bmatrix} ds_0 = extremely\ poor, ds_1 = very\ poor, \\ ds_2 = poor, ds_3 = medium, ds_4 = good, \\ ds_5 = very\ good, ds_6 = extremely\ good. \end{Bmatrix}$$

Definition 2[6-7]**.** The SVNSs is administrated:

$$d\eta = \left\{ \left(\theta, \phi_\eta(\theta), \varphi_\eta(\theta), \gamma_\eta(\theta) \right) \middle| \theta \in \Theta \right\} \quad (1)$$

where $\phi_\eta(\theta), \varphi_\eta(\theta), \gamma_\eta(\theta) \in [0,1]$ is truth-membership (TM), indeterminacy-membership (IM) and falsity-membership (FM), $0 \le \phi_\eta(\theta) + \varphi_\eta(\theta) + \gamma_\eta(\theta) \le 3$.

Definition 3[39]**.** Let $d\delta_j \ (j = 1, 2, 3, \ldots, k)$ be 2TLSs. If $d\delta = \left\langle \left(ds_t, d\xi \right), \left(ds_i, d\psi \right), \left(ds_f, d\zeta \right) \right\rangle$ is administrated for $ds_t, ds_i, ds_f \in ds$, $d\xi, d\psi, d\zeta \in \left[0, 0.5 \right)$, where $\left(ds_t, d\xi \right), \left(ds_i, d\psi \right), \left(ds_f, d\zeta \right)$ is administrated for TM, IM and FM by employing 2TLSs, the 2TLNSs is administrated:

$$d\delta_j = \left\langle \left(ds_{t_j}, d\xi_j \right), \left(ds_{i_j}, d\psi_j \right), \left(ds_{f_j}, d\zeta_j \right) \right\rangle \quad (2)$$

where

$0 \le \Delta^{-1}\left(ds_{t_j}, d\xi_j \right) \le \phi, 0 \le \Delta^{-1}\left(ds_{i_j}, d\psi_j \right) \le \phi,$

$0 \le \Delta^{-1}\left(ds_{f_j}, d\zeta_j \right) \le \phi \quad , \quad 0 \le \Delta^{-1}\left(ds_{t_j}, d\xi_j \right) + \Delta^{-1}\left(ds_{i_j}, d\psi_j \right) + \Delta^{-1}\left(ds_{f_j}, d\zeta_j \right) \le 3\phi$.

Definition 4[27]. Let $d\delta = \left\langle \left(ds_t, d\xi \right), \left(ds_i, d\psi \right), \left(ds_f, d\zeta \right) \right\rangle$. The uncertain score function (USF) and uncertain accuracy function (UAF) are administrated:

$$USF\left(d\delta \right) = \frac{\begin{pmatrix} 2\phi + \Delta^{-1}\left(ds_t, d\xi \right) \\ -\Delta^{-1}\left(ds_i, d\psi \right) - \Delta^{-1}\left(ds_f, d\zeta \right) \end{pmatrix}}{3\phi},$$

$$USF\left(d\delta \right) \in \left[0,1 \right] \quad (3)$$

$$UAF(d\delta) = \frac{\phi + \Delta^{-1}\left(ds_t, d\xi \right) - \Delta^{-1}\left(ds_f, d\zeta \right)}{2\phi},$$

$$UAF(d\delta) \in \left[0,1 \right] \quad (4)$$

Definition 5[27]. Let $d\delta_1 = \left\langle \left(ds_{t_1}, d\xi_1 \right), \left(ds_{i_1}, d\psi_1 \right), \left(ds_{f_1}, d\zeta_1 \right) \right\rangle$ and $d\delta_2 = \left\langle \left(ds_{t_2}, d\xi_2 \right), \left(ds_{i_2}, d\psi_2 \right), \left(ds_{f_2}, d\zeta_2 \right) \right\rangle$, then

(1) *if* $USF\left(d\delta_1 \right) \prec USF\left(d\delta_2 \right)$, $d\delta_1 \prec d\delta_2$;

(2) *if* $USF\left(d\delta_1 \right) = USF\left(d\delta_2 \right), UAF\left(d\delta_1 \right) \prec UAF\left(d\delta_2 \right)$, $d\delta_1 \prec d\delta_2$;

(3) *if* $USF\left(d\delta_1 \right) = USF\left(d\delta_2 \right), UAF\left(d\delta_1 \right) = UAF\left(d\delta_2 \right)$, $d\delta_1 = d\delta_2$;

Definition 6[27]. Let $d\delta_1 = \left\langle \left(ds_{t_1}, d\xi_1 \right), \left(ds_{i_1}, d\psi_1 \right), \left(ds_{f_1}, d\zeta_1 \right) \right\rangle$, $d\delta_2 = \left\langle \left(ds_{t_2}, d\xi_2 \right), \left(ds_{i_2}, d\psi_2 \right), \left(ds_{f_2}, d\zeta_2 \right) \right\rangle$ and $d\delta = \left\langle \left(ds_t, d\xi \right), \left(ds_i, d\psi \right), \left(ds_f, d\zeta \right) \right\rangle$ be three 2TLNNs, then

$$d\delta_1 \oplus d\delta_2 = \left\{ \begin{array}{l} \Delta\left(\phi\left(\dfrac{\Delta^{-1}\left(ds_{t_1}, d\xi_1\right)}{\phi} + \dfrac{\Delta^{-1}\left(ds_{t_2}, d\xi_2\right)}{\phi} - \dfrac{\Delta^{-1}\left(ds_{t_1}, d\xi_1\right)}{\phi} \cdot \dfrac{\Delta^{-1}\left(ds_{t_2}, d\xi_2\right)}{\phi}\right)\right), \\ \Delta\left(\phi\left(\dfrac{\Delta^{-1}\left(ds_{i_1}, d\psi_1\right)}{\phi} \cdot \dfrac{\Delta^{-1}\left(ds_{i_2}, d\psi_2\right)}{\phi}\right)\right), \Delta\left(\phi\left(\dfrac{\Delta^{-1}\left(ds_{f_1}, d\zeta_1\right)}{\phi} \cdot \dfrac{\Delta^{-1}\left(ds_{f_2}, d\zeta_2\right)}{\phi}\right)\right) \end{array} \right\}; \tag{1}$$

$$d\delta_1 \otimes d\delta_2 = \left\{ \begin{array}{l} \Delta\left(\phi\left(\dfrac{\Delta^{-1}\left(ds_{t_1}, d\xi_1\right)}{\phi} \cdot \dfrac{\Delta^{-1}\left(ds_{t_2}, d\xi_2\right)}{\phi}\right)\right), \\ \Delta\left(\phi\left(\dfrac{\Delta^{-1}\left(ds_{i_1}, d\psi_1\right)}{\phi} + \dfrac{\Delta^{-1}\left(ds_{i_2}, d\psi_2\right)}{\phi} - \dfrac{\Delta^{-1}\left(ds_{i_1}, d\psi_1\right)}{\phi} \cdot \dfrac{\Delta^{-1}\left(ds_{i_2}, d\psi_2\right)}{\phi}\right)\right), \\ \Delta\left(\phi\left(\dfrac{\Delta^{-1}\left(ds_{f_1}, d\zeta_1\right)}{\phi} + \dfrac{\Delta^{-1}\left(ds_{f_2}, d\zeta_2\right)}{\phi} - \dfrac{\Delta^{-1}\left(ds_{f_1}, d\zeta_1\right)}{\phi} \cdot \dfrac{\Delta^{-1}\left(ds_{f_2}, d\zeta_2\right)}{\phi}\right)\right) \end{array} \right\}; \tag{2}$$

$$\xi d\delta = \left\{\Delta\left(\phi\left(1-\left(1-\frac{\Delta^{-1}\left(ds_t, d\xi\right)}{\phi}\right)^{\xi}\right)\right), \Delta\left(\phi\left(\frac{\Delta^{-1}\left(ds_i, d\psi\right)}{\phi}\right)^{\xi}\right), \Delta\left(D\left(\frac{\Delta^{-1}\left(ds_f, d\zeta\right)}{\phi}\right)^{\xi}\right)\right\}, \xi > 0; \tag{3}$$

$$d\delta^{\xi} = \left\{\Delta\left(\phi\left(\frac{\Delta^{-1}\left(ds_t, d\xi\right)}{\phi}\right)^{\xi}\right), \Delta\left(\phi\left(1-\left(1-\frac{\Delta^{-1}\left(ds_i, d\psi\right)}{\phi}\right)^{\xi}\right)\right), \Delta\left(\phi\left(1-\left(1-\frac{\Delta^{-1}\left(ds_f, d\zeta\right)}{\phi}\right)^{\xi}\right)\right)\right\}, \xi > 0. \tag{4}$$

Definition 7[40]. Let $d\delta_1 = \left\langle \left(ds_{t_1}, d\xi_1\right), \left(ds_{i_1}, d\psi_1\right), \left(ds_{f_1}, d\zeta_1\right)\right\rangle$, $d\delta_2 = \left\langle \left(ds_{t_2}, d\xi_2\right), \left(ds_{i_2}, d\psi_2\right), \left(ds_{f_2}, d\zeta_2\right)\right\rangle$, then the 2TLNN Hamming distance (2TLNNHD) is constructed:

$$2TLNNHD\left(d\delta_1, d\delta_2\right) = \frac{1}{3}\left(\begin{array}{l} \left|\dfrac{\Delta^{-1}\left(ds_{t_1}, d\xi_1\right) - \Delta^{-1}\left(ds_{t_2}, d\xi_2\right)}{\phi}\right| \\ + \left|\dfrac{\Delta^{-1}\left(ds_{i_1}, d\psi_1\right) - \Delta^{-1}\left(ds_{i_2}, d\psi_2\right)}{\phi}\right| \\ + \left|\dfrac{\Delta^{-1}\left(ds_{f_1}, d\zeta_1\right) - \Delta^{-1}\left(ds_{f_2}, d\zeta_2\right)}{\phi}\right| \end{array}\right) \tag{5}$$

The 2TLNNWA approach is administrated:

Definition 7[39]. Let $d\delta_j = \left\langle \left(ds_{t_j}, d\xi_j\right), \left(ds_{i_j}, d\psi_j\right), \left(ds_{f_j}, d\zeta_j\right)\right\rangle$, the 2TLNNWA approach is administrated:

$$2\text{TLNNWA}\left(d\delta_1, d\delta_2, \ldots, d\delta_n\right) = \overset{n}{\underset{j=1}{\oplus}} dw_j d\delta_j = \left(\begin{array}{l} \Delta\left(\phi\left(1-\prod_{j=1}^{n}\left(1-\dfrac{\Delta^{-1}\left(ds_{t_j}, d\alpha_j\right)}{\phi}\right)^{w_j}\right)\right), \\ \Delta\left(\phi\prod_{j=1}^{n}\left(\dfrac{\Delta^{-1}\left(ds_{i_j}, d\beta_j\right)}{\phi}\right)^{w_j}\right), \\ \Delta\left(\phi\prod_{j=1}^{n}\left(\dfrac{\Delta^{-1}\left(ds_{f_j}, d\chi_j\right)}{\phi}\right)^{w_j}\right) \end{array}\right) \tag{6}$$

IV. 2TLNN-EXPTODIM-GRA APPROACH FOR MAGDM WITH ENTROPY

A. 2TLNN-MAGDM Data

The 2TLNN-ExpTODIM-GRA approach is administrated for MAGDM. Let $DA=\{DA_1,DA_2,\cdots,DA_m\}$ be alternatives, and $DG=\{DG_1,DG_2,\cdots,DG_n\}$ be attributes with weight numbers $dw=(dw_1,dw_2,\cdots,dw_n)$, where $dw_j\in[0,1]$, $\sum_{j=1}^{n}dw_j=1$ and experts $DE=\{DE_1,DE_2,\cdots,DE_q\}$ with weight numbers $d\omega=(d\omega_1,d\omega_2,\cdots,d\omega_q)$ $d\omega_j\in[0,1],\sum_{j=1}^{q}d\omega_j=1.$

Then, 2TLNN-ExpTODIM-GRA approach is administrated for MAGDM.

Step 1. Build the 2TLNN-matrix $DM=\left[DM_{ij}^{(t)}\right]_{m\times n}=\left\{\begin{matrix}\left(ds_{t_{ij}}^{(t)},d\alpha_{ij}^{(t)}\right),\\ \left(ds_{i_{ij}}^{(t)},d\beta_{ij}^{(t)}\right),\left(ds_{f_{ij}}^{(t)},d\chi_{ij}^{(t)}\right)\end{matrix}\right\}_{m\times n}$ and manage the average matrix $DM=\left[DM_{ij}\right]_{m\times n}$:

$$DM^{(t)}=\left[DM_{ij}^{(t)}\right]_{m\times n}=\begin{matrix}\\DA_1\\DA_2\\\vdots\\DA_m\end{matrix}\begin{matrix}\begin{matrix}DG_1&DG_2&\cdots&DG_n\end{matrix}\\\begin{bmatrix}DM_{11}^{(t)}&DM_{12}^{(t)}&\cdots&DM_{1n}^{(t)}\\DM_{21}^{(t)}&DM_{22}^{(t)}&\cdots&DM_{2n}^{(t)}\\\vdots&\vdots&\vdots&\vdots\\DM_{m1}^{(t)}&DM_{m2}^{(t)}&\cdots&DM_{mn}^{(t)}\end{bmatrix}\end{matrix} \quad (7)$$

$$DM=\left[DM_{ij}\right]_{m\times n}=\begin{matrix}\\DA_1\\DA_2\\\vdots\\DA_m\end{matrix}\begin{matrix}\begin{matrix}DG_1&DG_2&\cdots&DG_n\end{matrix}\\\begin{bmatrix}DM_{11}&DM_{12}&\cdots&DM_{1n}\\DM_{21}&DM_{22}&\cdots&DM_{2n}\\\vdots&\vdots&\vdots&\vdots\\DM_{m1}&DM_{m2}&\cdots&DM_{mn}\end{bmatrix}\end{matrix} \quad (8)$$

In line with the 2TLNNWA approach, the $DM=\left[DM_{ij}\right]_{m\times n}=\left\{\begin{matrix}\left(ds_{t_{ij}},d\alpha_{ij}\right),\\ \left(ds_{i_{ij}},d\beta_{ij}\right),\left(ds_{f_{ij}},d\chi_{ij}\right)\end{matrix}\right\}_{m\times n}$ is:

$$DM_{ij}=d\omega_1DM_{ij}^{1}\oplus d\omega_2DM_{ij}^{2}\oplus\cdots\oplus d\omega_tDM_{ij}^{t}$$

$$=\left\{\begin{matrix}\Delta\left(\phi\left(1-\prod_{t=1}^{q}\left(1-\frac{\Delta^{-1}\left(ds_{t_{ij}}^{(t)},d\alpha_{ij}^{(t)}\right)}{\phi}\right)^{dw_t}\right)\right),\\ \Delta\left(\phi\prod_{t=1}^{q}\left(\frac{\Delta^{-1}\left(ds_{i_{ij}}^{(t)},d\beta_{ij}^{(t)}\right)}{\phi}\right)^{dw_t}\right),\\ \Delta\left(\phi\prod_{t=1}^{q}\left(\frac{\Delta^{-1}\left(ds_{f_{ij}}^{(t)},d\chi_{ij}^{(t)}\right)}{\phi}\right)^{dw_t}\right)\end{matrix}\right\} \quad (9)$$

Step 2. Administrate the normalized $NDM=\left[NDM_{ij}\right]_{m\times n}=\left\{\begin{matrix}\left(ds_{t_{ij}},d\alpha_{ij}\right),\\ \left(ds_{i_{ij}},d\beta_{ij}\right),\left(ds_{f_{ij}},d\chi_{ij}\right)\end{matrix}\right\}_{m\times n}$ in light with $DM=\left[DM_{ij}\right]_{m\times n}=\left\{\begin{matrix}\left(ds_{t_{ij}},d\alpha_{ij}\right),\\ \left(ds_{i_{ij}},d\beta_{ij}\right),\left(ds_{f_{ij}},d\chi_{ij}\right)\end{matrix}\right\}_{m\times n}$.

For beneficial attributes:

$$NDM_{ij}=DM_{ij}=\left\{\begin{matrix}\left(ds_{t_{ij}},d\alpha_{ij}\right),\\ \left(ds_{i_{ij}},d\beta_{ij}\right),\left(ds_{f_{ij}},d\chi_{ij}\right)\end{matrix}\right\} \quad (10)$$

For non-beneficial attributes:

$$NDM_{ij}=\left\{\left(ds_{t_{ij}},d\alpha_{ij}\right),\left(ds_{i_{ij}},d\beta_{ij}\right),\left(ds_{f_{ij}},d\chi_{ij}\right)\right\}$$

$$=\left\{\begin{matrix}\Delta\left(\phi-\Delta^{-1}\left(ds_{t_{ij}},d\alpha_{ij}\right)\right),\\ \Delta\left(\phi-\Delta^{-1}\left(ds_{i_{ij}},d\beta_{ij}\right)\right),\Delta\left(\phi-\Delta^{-1}\left(ds_{f_{ij}},d\chi_{ij}\right)\right)\end{matrix}\right\} \quad (11)$$

B. Administrate the Weight Numbers with Entropy

Step 3. Normalized hybrid decision matrix $NHDM_{ij}$ is administrated through Entropy [41]:

$$NHDM_{ij} = \frac{\dfrac{UAF\left\{\left(ds_{t_{ij}}, d\alpha_{ij}\right),\left(ds_{i_{ij}}, d\beta_{ij}\right),\left(ds_{f_{ij}}, d\chi_{ij}\right)\right\}}{USF\left\{\left(ds_{t_{ij}}, d\alpha_{ij}\right),\left(ds_{i_{ij}}, d\beta_{ij}\right),\left(ds_{f_{ij}}, d\chi_{ij}\right)\right\}}}{\sum_{i=1}^{m}\left(\dfrac{UAF\left\{\left(ds_{t_{ij}}, d\alpha_{ij}\right),\left(ds_{i_{ij}}, d\beta_{ij}\right),\left(ds_{f_{ij}}, d\chi_{ij}\right)\right\}}{USF\left\{\left(ds_{t_{ij}}, d\alpha_{ij}\right),\left(ds_{i_{ij}}, d\beta_{ij}\right),\left(ds_{f_{ij}}, d\chi_{ij}\right)\right\}}\right)}, \tag{12}$$

Then, the uncertain hybrid information Shannon entropy (UHISE) is administrated:

$$UHISE_j = -\frac{1}{\ln m}\sum_{i=1}^{m} NHDM_{ij} \ln NHDM_{ij} \tag{13}$$

and $NHDM_{ij} \ln NHDM_{ij} = 0$ if $NHDM_{ij} = 0$.

Then, the weights values $dw = \left(dw_1, dw_2, \cdots, dw_n\right)$ is managed:

$$dw_j = \frac{1 - UHISE_j}{\sum_{j=1}^{n}\left(1 - UHISE_j\right)} \tag{14}$$

C. 2TLNN-ExpTODIM-GRA Approach for MAGDM

2TLNN-ExpTODIM-GRA approach is administrated for MAGDM.

Step 4. Administrate relative weight:

$$rdw_j = dw_j \Big/ \max_j dw_j, \tag{15}$$

Step 5. 2TLNN dominance degree (2TLNNDD) of DA_i for DA_t under DG_j is administrated:

$$2TLNNDD_j\left(DA_i, DA_t\right) = \begin{cases} \dfrac{rdw_j \times \left(1 - 10^{-\rho 2TLNNHDHD\left(NDM_{ij}, NDM_{tj}\right)}\right)}{\sum_{j=1}^{n} rdw_j} & \text{if } SF\left(NDM_{ij}\right) > SF\left(NDM_{tj}\right) \\ 0 & \text{if } SF\left(NDM_{ij}\right) = SF\left(NDM_{tj}\right) \\ -\dfrac{1}{\theta}\dfrac{\sum_{j=1}^{n} rdw_j \times \left(1 - 10^{-\rho 2TLNNHDHD\left(NDM_{ij}, NDM_{tj}\right)}\right)}{rdw_j} & \text{if } SF\left(NDM_{ij}\right) < SF\left(NDM_{tj}\right) \end{cases} \tag{16}$$

where θ is administrated [42] and $\rho \in [1,5]$ [36].

The 2TLNNDD under DG_j is administrated:

$$2TLNNDD_j\left(DA_i\right) = \left[2TLNNDD_j\left(DA_i, DA_t\right)\right]_{m\times m}$$

$$= \begin{array}{c} \\ DA_1 \\ DA_2 \\ \vdots \\ DA_m \end{array} \begin{array}{c} \begin{array}{cccc} DA_1 & DA_2 & \cdots & DA_m \end{array} \\ \begin{bmatrix} 0 & 2TLNNDD_j\left(DA_1, DA_2\right) & \cdots & 2TLNNDD_j\left(DA_1, DA_m\right) \\ 2TLNNDD_j\left(DA_2, DA_1\right) & 0 & \cdots & 2TLNNDD_j\left(DA_2, DA_m\right) \\ \vdots & \vdots & \cdots & \vdots \\ 2TLNNDD_j\left(DA_m, DA_1\right) & 2TLNNDD_j\left(DA_m, DA_2\right) & \cdots & 0 \end{bmatrix} \end{array}$$

(3) Administrate the overall 2TLNNDD of DA_i for other alternatives under DG_j:

$$2TLNNDD_j\left(DA_i\right) = \sum_{t=1}^{m} 2TLNNDD_j\left(DA_i, DA_t\right) \tag{17}$$

The overall 2TLNNDD is administrated:

$$2TLNNDD=\left(2TLNNDD_{ij}\right)_{m\times n}$$

$$=\begin{bmatrix} & DG_1 & DG_2 & \cdots & DG_n \\ DA_1 & \sum_{t=1}^{m}2TLNNDD_1\left(DA_1,DA_t\right) & \sum_{t=1}^{m}2TLNNDD_2\left(DA_1,DA_t\right) & \cdots & \sum_{t=1}^{m}2TLNNDD_n\left(DA_1,DA_t\right) \\ DA_2 & \sum_{t=1}^{m}2TLNNDD_1\left(DA_2,DA_t\right) & \sum_{t=1}^{m}2TLNNDD_2\left(DA_2,DA_t\right) & \cdots & \sum_{t=1}^{m}2TLNNDD_n\left(DA_2,DA_t\right) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ DA_m & \sum_{t=1}^{m}2TLNNDD_1\left(DA_m,DA_t\right) & \sum_{t=1}^{m}2TLNNDD_2\left(DA_m,DA_t\right) & \cdots & \sum_{t=1}^{m}2TLNNDD_n\left(DA_m,DA_t\right) \end{bmatrix} \quad (18)$$

Step 6. Administrate the 2TLNNPIDS (2TLNN positive ideal decision solution) and 2TLNNNIDS (2TLNN negative ideal decision solution):

$$2TLNNPIDS=\left(2TLNNPIDS_1,2TLNNPIDS_1,\cdots,2TLNNPIDS_n\right) \quad (19)$$

$$2TLNNNIDS=\left(2TLNNNIDS_1,2TLNNNIDS_1,\cdots,2TLNNNIDS_n\right) \quad (20)$$

$$2TLNNPIDS_j=\max_i\left(2TLNNDD_{ij}\right) \quad (21\text{-a})$$

$$2TLNNNIDS_j=\min_i\left(2TLNNDD_{ij}\right) \quad (21\text{-b})$$

Step 7. Administrate the 2TLNNGRC (2TLNN grey rational coefficients) from 2TLNNPIDS and 2TLNNNIDS.

$$2TLNNGRC\left(\zeta_{ij}^{+}\right)=\frac{\left(\begin{array}{l}\min\limits_{1\le i\le m}\min\limits_{1\le j\le n}\left|2TLNNDD_{ij}-2TLNNPIDS_j\right| \\ +\rho\max\limits_{1\le i\le m}\max\limits_{1\le j\le n}\left|2TLNNDD_{ij}-2TLNNPIDS_j\right|\end{array}\right)}{\left(\begin{array}{l}\left|2TLNNDD_{ij}-2TLNNPIDS_j\right| \\ +\rho\max\limits_{1\le i\le m}\max\limits_{1\le j\le n}\left|2TLNNDD_{ij}-2TLNNPIDS_j\right|\end{array}\right)} \quad (22)$$

$$2TLNNGRC\left(\zeta_{ij}^{-}\right)=\frac{\left(\begin{array}{l}\min\limits_{1\le i\le m}\min\limits_{1\le j\le n}\left|2TLNNDD_{ij}-2TLNNNIDS_j\right| \\ +\rho\max\limits_{1\le i\le m}\max\limits_{1\le j\le n}\left|2TLNNDD_{ij}-2TLNNNIDS_j\right|\end{array}\right)}{\left(\begin{array}{l}\left|2TLNNDD_{ij}-2TLNNNIDS_j\right| \\ +\rho\max\limits_{1\le i\le m}\max\limits_{1\le j\le n}\left|2TLNNDD_{ij}-2TLNNNIDS_j\right|\end{array}\right)} \quad (23)$$

Step 8. Administrate the 2TLNNGRD (2TLNN grey relation degree) from 2TLNNPIDS and 2TLNNNIDS.

$$2TLNNGRD\left(\zeta_i^{+}\right)=\sum_{j=1}^{n}dw_j\,2TLNNGRC\left(\zeta_{ij}^{+}\right) \quad (24)$$

$$2TLNNGRD\left(\zeta_i^{-}\right)=\sum_{j=1}^{n}dw_j\,2TLNNGRC\left(\zeta_{ij}^{-}\right) \quad (25)$$

Step 9. Administrate the 2TLNNRRD (2TLNN relative relational degree) from the 2TLNNPIDS.

$$2TLNNRRD\left(\zeta_i\right)=\frac{2TLNNGRD\left(\zeta_i^{+}\right)}{2TLNNGRD\left(\zeta_i^{-}\right)+2TLNNGRD\left(\zeta_i^{+}\right)}=\frac{\sum_{j=1}^{n}dw_j\,2TLNNGRC\left(\zeta_{ij}^{+}\right)}{\sum_{j=1}^{n}dw_j\,2TLNNGRC\left(\zeta_{ij}^{+}\right)+\sum_{j=1}^{n}dw_j\,2TLNNGRC\left(\zeta_{ij}^{-}\right)} \quad (26)$$

Step 10. Sort and choose the optimal alternative in line with largest 2TLNNRRD.

V. Numerical Example and Comparative Analysis

A. Numerical Example

Under the background of rural revitalization strategy and development of rural economy and society in the new era, construction of quality evaluation for PIE in rural areas needs to break free from the shackles of traditional concepts in terms of evaluation standards, evaluation orientation, evaluation mechanisms, evaluation methods, and more. It requires establishing a distinctive view of quality and a new perspective on evaluation, focusing on student-centered and comprehensive development, public governance and collaborative education, diagnostic improvement, and giving importance to the process. It is necessary to prioritize laws and policies, base standards on the highest level, emphasize individual characteristics, and prioritize systematic norms. The primary focus lies in the core elements and essential links of integrated vocational education and industry development, which aim to lead and promote deep

integration and high-quality advancement of PEI in rural areas, thereby contributing to rural revitalization. The PEI quality evaluation in RVE under rural revitalization falls under the framework of MAGDM. In this evaluation, there are five potential vocational and technical colleges $DA_i\left(i=1,2,3,4,5\right)$ to choose through four attributes [43]: DG1: Social benefit for PEI. DG2: management cost for PEI. DG3: Economic benefits for PEI. DG4: Production-education integration achievements. Five vocational and technical colleges $DA_i\left(i=1,2,3,4,5\right)$ are assessed using linguistic scales (refer to Table I [40]) under four criteria. The evaluation involves three experts $DE_t\left(t=1,2,3\right)$, each with their respective weight $d\omega=\left(1/3,1/3,1/3\right)$ in the assessment process.

TABLE I. LINGUISTIC INFORMATION SCALES AND 2TLNNS

Linguistic Terms	2TLNNs
Exceedingly Terrible-DET	$\{(ds_0,0),\ (ds_5,0),\ (ds_6,0)\}$
Very Terrible-DVT	$\{(ds_1,0),\ (ds_4,0),\ (ds_5,0)\}$
Terrible-DT	$\{(ds_2,0),\ (ds_3,0),\ (ds_4,0)\}$
Medium-DM	$\{(ds_3,0),\ (ds_3,0),\ (ds_3,0)\}$
Well-DW	$\{(ds_4,0),\ (ds_3,0),\ (ds_2,0)\}$
Very Well-DVW	$\{(ds_5,0),\ (ds_2,0),\ (ds_1,0)\}$
Exceedingly Well-DEW	$\{(ds_6,0),\ (ds_1,0),\ (ds_0,0)\}$

The PEI quality evaluation in RVE under rural revitalization is managed using the 2TLNN-ExpTODIM-GRA approach.

Step 1. Construct the 2TLNN-matrix $DM=\left[DM_{ij}^{(t)}\right]_{5\times4}\left(t=1,2,3\right)$ (Table II - IV).

Then, from 2TLNNWA approach, the $DM=\left[DM_{ij}\right]_{5\times4}$ is constructed (Table V).

Step 2. Normalize the $DM=\left[DM_{ij}\right]_{5\times4}$ into $NDM=\left[NDM_{ij}\right]_{5\times4}$ (see Table VI).

Step 3. Administrate the weight (Table VII):

Step 4. Administrate the relative weight (see Table VIII):

Step 5. Administrate the $2TLNNDD=\left(2TLNNDD_{ij}\right)_{5\times4}$ (Table IX):

Step 6. Administrate the 2TLNNPIDS and 2TLNNNIDS (Table X).

Step 7. Administrate the $2TLNNGRC\left(\zeta_{ij}^{+}\right)$ and $2TLNNGRC\left(\zeta_{ij}^{-}\right)$ (Table XI-XII).

Step 8. Establish the $2TLNNGRD\left(\zeta_{i}^{+}\right)$ and $2TLNNGRD\left(\zeta_{i}^{-}\right)$ (Table XIII).

Step 9. Establish the 2TLNNRRD (Table XIV).

TABLE II. DECISION DATA FROM DE_1

	DG_1	DG_2	DG_3	DG_4
DA_1	DM	DW	DT	DM
DA_2	DW	DVW	DT	DM
DA_3	DVT	DVT	DM	DW
DA_4	DM	DT	DVW	DW
DA_5	DW	DVW	DVT	DT

TABLE III. DECISION DATA FROM DE_2

	DG_1	DG_2	DG_3	DG_4
DA_1	DVW	DT	DW	DM
DA_2	DM	DVW	DW	DVT
DA_3	DVW	DM	DW	DVT
DA_4	DW	DM	DT	DVW
DA_5	DT	DVT	DM	DVW

TABLE IV. DECISION DATA FROM DE_3

	DG_1	DG_2	DG_3	DG_4
DA_1	DVT	DVT	DVW	DM
DA_2	DT	DVW	DVT	DM
DA_3	DVW	DW	DM	DVT
DA_4	DW	DM	DT	DVW
DA_5	DM	DW	DVW	DVT

TABLE V. THE $DM = \left[DM_{ij} \right]_{5\times4}$

	DG_1
DA_1	$\{(ds_4,-0.1294),(ds_3,-0.4493),(ds_2,0.1175)\}$
DA_2	$\{(ds_4,-0.1576),(ds_3,-0.3438),(ds_2,0.1072)\}$
DA_3	$\{(ds_3,0.3672),(ds_3,0.2714),(ds_3,-0.3098)\}$
DA_4	$\{(ds_3,0.1296),(ds_3,0.2128),(ds_3,-0.0394)\}$
DA_5	$\{(ds_4,-0.4564),(ds_3,-0.2296),(ds_2,0.4565)\}$
	DG_2
DA_1	$\{(ds_2,-0.2769),(ds_3,0.2714),(ds_4,0.2335)\}$
DA_2	$\{(ds_4,0.1435),(ds_3,0.1342),(ds_2,0.1657)\}$
DA_3	$\{(ds_4,-0.3523),(ds_3,-0.3436),(ds_2,0.3127)\}$
DA_4	$\{(ds_5,-0.3195),(ds_2,0.3524),(ds_1,0.3139)\}$
DA_5	$\{(ds_4,0.1654),(ds_3,-0.3436),(ds_2,-0.1763)\}$
	DG_3
DA_1	$\{(ds_2,-0.0327),(ds_3,0.3659),(ds_4,0.1298)\}$
DA_2	$\{(ds_1,0.3457),(ds_4,-0.3523),(ds_5,-0.3452)\}$
DA_3	$\{(ds_4,-0.1572),(ds_3,-0.3438),(ds_2,0.1576)\}$
DA_4	$\{(ds_4,0.3793),(ds_2,0.4625),(ds_2,-0.2697)\}$
DA_5	$\{(ds_3,-0.4968),(ds_3,0.2714),(ds_3,0.4975)\}$
	DG_4
DA_1	$\{(ds_3,-0.4019),(ds_3,0.2762),(ds_3,0.4954)\}$

DA_2	$\{(ds_5,-0.3914),(ds_2,0.2589),(ds_1,0.3207)\}$
DA_3	$\{(ds_3,-0.2584),(ds_3,0.3655),(ds_3,0.2582)\}$
DA_4	$\{(cs_3,-0.4743),(cs_3,0.2715),(hs_3,0.4746)\}$
DA_5	$\{(ds_4,0.4843),(ds_3,-0.4492),(ds_2,-0.4102)\}$

$$NDM=\left[NDM_{ij}\right]_{5\times 4}$$

TABLE VI. THE

	DG_1
DA_1	$\{(ds_4,-0.1294),(ds_3,-0.4493),(ds_2,0.1175)\}$
DA_2	$\{(ds_4,-0.1576),(ds_3,-0.3438),(ds_2,0.1072)\}$
DA_3	$\{(ds_3,0.3672),(ds_3,0.2714),(ds_3,-0.3098)\}$
DA_4	$\{(ds_3,0.1296),(ds_3,0.2128),(ds_3,-0.0394)\}$
DA_5	$\{(ds_4,-0.4564),(ds_3,-0.2296),(ds_2,0.4565)\}$
	DG_2
DA_1	$\{(ds_4,0.2769),(ds_3,-0.2714),(ds_2,-0.2335)\}$
DA_2	$\{(ds_2,-0.1435),(ds_3,-0.1342),(ds_4,-0.1657)\}$
DA_3	$\{(ds_2,0.3523),(ds_3,0.3436),(ds_4,-0.3127)\}$
DA_4	$\{(ds_1,0.3195),(ds_4,-0.3524),(ds_5,-0.3139)\}$
DA_5	$\{(ds_2,-0.1654),(ds_3,0.3436),(ds_4,0.1763)\}$
	DG_3
DA_1	$\{(ds_2,-0.0327),(ds_3,0.3659),(ds_4,0.1298)\}$
DA_2	$\{(ds_1,0.3457),(ds_4,-0.3523),(ds_5,-0.3452)\}$
DA_3	$\{(ds_4,-0.1572),(ds_3,-0.3438),(ds_2,0.1576)\}$
DA_4	$\{(ds_4,0.3793),(ds_2,0.4625),(ds_2,-0.2697)\}$
DA_5	$\{(ds_3,-0.4968),(ds_3,0.2714),(ds_3,0.4975)\}$
	DG_4
DA_1	$\{(ds_3,-0.4019),(ds_3,0.2762),(ds_3,0.4954)\}$
DA_2	$\{(ds_5,-0.3914),(ds_2,0.2589),(ds_1,0.3207)\}$
DA_3	$\{(ds_3,-0.2584),(ds_3,0.3655),(ds_3,0.2582)\}$
DA_4	$\{(cs_3,-0.4743),(cs_3,0.2715),(hs_3,0.4746)\}$
DA_5	$\{(ds_4,0.4843),(ds_3,-0.4492),(ds_2,-0.4102)\}$

TABLE VII. WEIGHT NUMBERS

	DG_1	DG_2	DG_3	DG_4
weight	0.1705	0.2726	0.2806	0.2763

TABLE VIII. RELATIVE WEIGHT

	DG_1	DG_2	DG_3	DG_4
relative weight	0.6076	0.9715	1.0000	0.9847

$$2TLNNDD = \left(2TLNNDD_{ij}\right)_{5\times4}$$

TABLE IX.

	DG_1	DG_2	DG_3	DG_4
DA_1	-0.2568	1.1397	0.4547	-0.8459
DA_2	0.2467	0.3815	-2.5416	0.6480
DA_3	-1.1506	-0.9523	1.1357	-1.8215
DA_4	0.7805	-0.5346	-0.6770	1.4450
DA_5	-1.0441	-1.5122	-0.8070	-0.0889

TABLE X. 2TLNNPIDS AND 2TLNNNIDS

	DG_1	DG_2	DG_3	DG_4
2TLNNPIDS	0.7805	1.1397	1.1357	1.4450
2TLNNNIDS	-1.1506	-1.5122	-2.5416	-1.8215

$$2TLNNGRC\left(\zeta_{ij}^{+}\right)$$

TABLE XI.

	DG_1	DG_2	DG_3	DG_4
DA_1	0.6393	1.0000	0.7297	0.4452
DA_2	0.7750	0.7080	0.3333	0.6976
DA_3	0.4877	0.4678	1.0000	0.3602
DA_4	1.0000	0.5234	0.5035	1.0000
DA_5	0.5019	0.4095	0.4862	0.5452

$$2TLNNGRC\left(\zeta_{ij}^{-}\right)$$

TABLE XII.

	DG_1	DG_2	DG_3	DG_4
DA_1	0.6729	0.4095	0.3803	0.6534
DA_2	0.5682	0.4926	1.0000	0.4268
DA_3	1.0000	0.7666	0.3333	1.0000
DA_4	0.4877	0.6529	0.4965	0.3602
DA_5	0.9453	1.0000	0.5146	0.5149

$$2TLNNGRD\left(\zeta_{i}^{+}\right) \text{ AND } 2TLNNGRD\left(\zeta_{i}^{-}\right)$$

TABLE XIII. THE

Alternative	$2TLNNGRD\left(\zeta_{i}^{+}\right)$	$2TLNNGRD\left(\zeta_{i}^{-}\right)$
DA_1	0.7094	0.5136
DA_2	0.6114	0.6297
DA_3	0.5908	0.7493
DA_4	0.7308	0.5000
DA_5	0.4843	0.7204

TABLE XIV. 2TLNNRRD AND ORDER

Alternative	2TLNNRRD	Order
DA_1	0.5801	2
DA_2	0.4926	3
DA_3	0.4409	4
DA_4	0.5938	1
DA_5	0.4020	5

Thus, the order is administrated: $DA_4 > DA_1 > DA_2 > DA_3 > DA_5$ and the optimal vocational and technical college is DA_4.

B. Comparative Analysis

Then, 2TLNN-ExpTODIM-GRA approach is fully compared with 2TLNNWA approach[39], 2TLNNWG approach [39], 2TLNN-CODAS approach [44], 2TLNN-GRA approach [45], 2TLNN-CLVA approach [46] and 2TLNN-TODIM approach [40]. The comparative result is administrated in Table XV.

TABLE XV. ORDER FOR DIFFERENT APPROACHES

Approaches	Order
2TLNNWA approach [39]	$DA_4 > DA_1 > DA_2 > DA_3 > DA_5$
2TLNNWG approach [39]	$DA_4 > DA_1 > DA_3 > DA_2 > DA_5$
2TLNN-CODAS approach [44]	$DA_4 > DA_1 > DA_2 > DA_3 > DA_5$
2TLNN-GRA approach [45]	$DA_4 > DA_1 > DA_2 > DA_3 > DA_5$
2TLNN-TODIM approach [40]	$DA_4 > DA_1 > DA_2 > DA_3 > DA_5$

The comparative analysis conducted in this study confirms the reasonableness and effectiveness of the 2TLNN-ExpTODIM-GRA approach. However, it is worth highlighting that main limitations of 2TLNN-ExpTODIM-GRA approach are the lack of emphasis on consensus for MAGDM.

VI. CONCLUSION

The rural revitalization strategy is significant decision made by Party and the state, guided by the ideology of socialism with Chinese characteristics. Within this strategy, RVE serves as a crucial force in promoting self-realization among adults, breaking the cycle of poverty across generations, fostering rural industrial development, and advancing the modernization of agriculture and rural areas. Despite its importance, RVE in our country faces practical challenges such as the need to align with the accelerating agricultural modernization, the underutilization of its functional role, the transition of the rural population into employment and the corresponding lack of momentum in its development, the persistence of the "jumping off the farm gate" mindset and its deviation from the goals of RVE training, and the increasing pace of rural population aging with a delayed response from RVE. The PEI quality evaluation in RVE under the rural revitalization context involves MAGDM. In recent times, ExpTODIM approach and GRA approach was employed to manage MAGDM. To address the PEI quality evaluation in RVE under rural revitalization, the utilization of 2TLNSs proves valuable for representing uncertain information. In this study, we propose the 2TLNN-ExpTODIM-GRA approach to effectively handle MAGDM within the framework of 2TLNSs. Additionally, a numerical study is conducted to validate the implementation of this approach for the PEI quality evaluation in RVE under rural revitalization. The primary contribution of this paper lies in the development of ExpTODIM-GRA approach, specifically tailored to address MAGDM within the context of 2TLNSs, with a specific focus on the PEI quality evaluation in RVE under rural revitalization.

ACKNOWLEDGMENT

This work was supported by The 2023 Discipline Co construction Project of Guangdong Provincial Philosophy and Social Sciences Plan, research on The Mechanism of Vocational Education's Anti-Poverty Action and Guangdong's Implementation Path under Project No. GD23XJY91.

REFERENCES

[1] P. Yan, H. Yan, L. Ma, Research on model and mechanism of higher vocational education in the rural, in: 2009 International Conference on Information Management, Innovation Management and Industrial Engineering, ICIII 2009, December 26, 2009 - December 27, 2009, IEEE Computer Society, Xi'an, China, 2009, pp. 214-217.

[2] B. Bhavani, S. Sheshadri, R. Unnikrishnan, Vocational education technology: Rural India, in: Proceedings of the 1st Amrita ACM-W Celebration of Women in Computing in India, A2CWiC'10, Association for Computing Machinery, 2010.

[3] Y. Ning, Q. Han, S. Zhang, S. Jie, C. Pei, Teaching reform of rural vocational education based on modern educational technology, in: ICEIT 2010 - 2010 International Conference on Educational and Information Technology, Proceedings, IEEE Computer Society, 2010, pp. V3435-V3438.

[4] C. Chu, Q. Jing, The comprehensive evaluation of vocational education quality in rural human resource development, in: International Conference on Management and Service Science, MASS 2011, IEEE Computer Society, 2011.

[5] C. Weng, R. Yan, Situation analysis and countermeasure study on rural

vocational education and training for modern farmers' cultivation based on information technology, in: Advances in Computer Science, Environment, Ecoinformatics, and Education - International Conference, CSEE 2011, Proceedings, Springer Verlag, 2011, pp. 76-80.

[6] N. Akshay, K. Sreeram, A. Anand, R. Venkataraman, R.R. Bhavani, MoVE: Mobile vocational education for rural India, in: 2012 IEEE International Conference on Technology Enhanced Education, ICTEE 2012, January 3, 2012 - January 5, 2012, IEEE Computer Society, Kerala, India, 2012.

[7] F. Wang, Study of vocational technical education in promotion of Chinese rural labor transfer, in: International Conference on Informatics and Management Science, IMS 2012, December 21, 2012 - December 23, 2012, Springer Verlag, Kunming, China, 2013, pp. 553-558.

[8] H. Liu, S.T. Liu, G.-L. Long, Vocational education of girls in rural areas analysis—Based in Hebei, in: 3rd International Conference on Key Engineering Materials and Computer Science, KEMCS 2014, August 5, 2014 - August 6, 2014, Trans Tech Publications Ltd, Singapore, Singapore, 2014, pp. 266-271.

[9] L. Sun, A study on higher vocational college libraries' serving for new rural vocational education, in: International Conference on Information Technology and Computer Application Engineering, ITCAE 2014, August 11, 2014 - August 12, 2014, CRC Press/Balkema, Hong Kong, China, 2015, pp. 747-750.

[10] F. Tian, Analysis of the Practical Path of Vocational Education Serving Rural Revitalization by Digital Means and Philosophy of Technology, Applied Mathematics and Nonlinear Sciences, 9 (2024).

[11] Y. Yang, The cultivation path of vocational talents in rural revitalization industry based on multidimensional data analysis method in the context of industry-education integration, Applied Mathematics and Nonlinear Sciences, 9 (2024).

[12] S. Mingfang, Effect of rural vocational education from the perspective of urbanization, Agro Food Ind. Hi-Tech, 28 (2017) 3338-3341.

[13] K.P. Sachith, A. Gopal, A. Muir, R.R. Bhavani, Contextualizing ICT Based Vocational Education for Rural Communities: Addressing Ethnographic Issues and Assessing Design Principles, in: 16th IFIP TC13 International Conference on Human–Computer Interaction, INTERACT 2017, September 25, 2017 - September 29, 2017, Springer Verlag, Mumbai, India, 2017, pp. 3-12.

[14] H. Wang, The Model Construction of Computer Network Technology in Vocational Education under the Rural Revitalization Strategy, in: 2nd International Conference on Information Science and Education, ICISE-IE 2021, November 26, 2021 - November 28, 2021, Institute of Electrical and Electronics Engineers Inc., Virtual, Nanchang, China, 2021, pp. 46-49.

[15] S. Huang, H. Chen, Research on quality evaluation of industry-education integration for rural vocational education in the perspective of rural revitalization with PL-MACONT method, Journal of Intelligent and Fuzzy Systems, 44 (2023) 9743-9755.

[16] B. Xiao, X. Ma, Research on the influencing factors of the precise supply of rural vocational education under the strategy of rural revitalization, in: 15th International Conference on Education Technology and Computers, ICETC 2023, September 26, 2023 - September 28, 2023, Association for Computing Machinery, Barcelona, Spain, 2023, pp. 364-369.

[17] B. Xiao, X. Wu, Z. Lin, X. Ma, H. Huang, Providing Targeted Rural Vocational Education to Serve Rural Revitalization from a Systematic Perspective, Computer-Aided Design and Applications, 20 (2023) 245-260.

[18] X. Juan, Evaluation on the Effect of Higher Vocational Education Empowering Rural Revitalization in Hunan Province, in: 13th International Conference on Logistics and Systems Engineering, March 29, 2024 - March 30, 2024, Aussino Academic Publishing House, Wuhan, China, 2024, pp. 168-176.

[19] A. Mondal, S.K. Roy, D. Pamucar, Regret-based three-way decision making with possibility dominance and SPA theory in incomplete information system, Expert Systems with Applications, 211 (2023) 17.

[20] D. Pamucar, G. Duran-Romero, M. Yazdani, A.M. Lopez, A decision analysis model for smart mobility system development under circular economy approach, Socio-Economic Planning Sciences, 86 (2023) 24.

[21] D. Pamucar, I. Gokasar, A.E. Torkayesh, M. Deveci, L. Martinez, Q. Wu, Prioritization of unmanned aerial vehicles in transportation systems using the integrated stratified fuzzy rough decision-making approach with the hamacher operator, Information Sciences, 622 (2023) 374-404.

[22] D. Pamucar, A.E. Torkayesh, S. Biswas, Supplier selection in healthcare supply chain management during the COVID-19 pandemic: a novel fuzzy rough decision-making approach, Annals of Operations Research, 328 (2023) 977-1019.

[23] H. Garg, K. Ullah, T. Mahmood, Z. Ali, H. Khalifa, Multi-attribute decision-making problems based on aggregation operators with complex interval-valued T-spherical fuzzy information, Maejo Int. J. Sci. Technol., 16 (2022) 51-65.

[24] H. Garg, Z. Ali, T. Mahmood, M.R. Ali, TOPSIS-method based on generalized dice similarity measures with hamy mean operators and its application to decision-making process, Alexandria Engineering Journal, 65 (2023) 383-397.

[25] H. Garg, Z. Ali, T. Mahmood, M.R. Ali, A. Alburaikan, Schweizer-Sklar prioritized aggregation operators for intuitionistic fuzzy information and their application in multi-attribute decision-making, Alexandria Engineering Journal, 67 (2023) 229-240.

[26] H. Garg, K. Ullah, K. Ali, M. Akram, M.N. Abid, Multi-attribute decision-making based on sine trigonometric aggregation operators for T-spherical fuzzy information, Soft Computing, (2023) 15.

[27] F. Xu, A 2TLNS-based exponential TODIM-EDAS approach for evaluating sustainable development of cross-border e-commerce platforms under uncertainty, Journal of Intelligent & Fuzzy Systems, 46 (2024) 6383-6398.

[28] S. Aydin, C. Kahraman, A spherical fuzzy multi expert MCDM method based on the entropy and cosine similarity, in: 15th Symposium of Intelligent Systems and Knowledge Engineering (ISKE) held jointly with 14th International FLINS Conference (FLINS), World Scientific Publ Co Pte Ltd, Cologne, GERMANY, 2020, pp. 157-164.

[29] C. Kahraman, Decision making using intelligent and fuzzy techniques Preface, Journal of Intelligent & Fuzzy Systems, 39 (2020) 5079-5079.

[30] N. Tuysuz, C. Kahraman, CODAS method using Z-fuzzy numbers, Journal of Intelligent & Fuzzy Systems, 38 (2020) 1649-1662.

[31] C. Kahraman, FUZZY DECISION MAKING: METHODOLOGIES AND APPLICATIONS Preface, Journal of Multiple-Valued Logic and Soft Computing, 37 (2021) 207-209.

[32] J. Ye, S.G. Du, R. Yong, Some aggregation operators of credibility trapezoidal fuzzy neutrosophic values and their decision-making application in the selection of slope design schemes, Journal of Intelligent & Fuzzy Systems, 43 (2022) 2803-2817.

[33] J. Ye, J.M. Song, S.G. Du, Correlation Coefficients of Consistency Neutrosophic Sets Regarding Neutrosophic Multi-valued Sets and Their Multi-attribute Decision-Making Method, International Journal of Fuzzy Systems, 24 (2022) 925-932.

[34] J. Ye, S.G. Du, R. Yong, Multi-criteria decision-making model using trigonometric aggregation operators of single-valued neutrosophic credibility numbers, Information Sciences, 644 (2023) 17.

[35] J. Ye, B.Z. Sun, X.L. Chu, J.M. Zhan, J.X. Cai, Valued outranking relation-based heterogeneous multi-decision multigranulation probabilistic rough set and its use in medical decision-making, Expert Systems with Applications, 228 (2023) 18.

[36] A.B. Leoneti, L. Gomes, A novel version of the TODIM method based on the exponential model of prospect theory: The ExpTODIM method, European Journal of Operational Research, 295 (2021) 1042-1055.

[37] H. Sun, Z. Yang, Q. Cai, G.W. Wei, Z.W. Mo, An extended Exp-TODIM method for multiple attribute decision making based on the Z-Wasserstein distance, Expert Systems with Applications, 214 (2023) 14.

[38] J. Deng, Introduction to grey system theory, The Journal of Grey System, 1 (1989) 1-24.

[39] J. Wang, G.W. Wei, Y. Wei, Models for Green Supplier Selection with Some 2-Tuple Linguistic Neutrosophic Number Bonferroni Mean Operators, Symmetry-Basel, 10 (2018) 36.

[40] J. Wang, G.W. Wei, M. Lu, TODIM Method for Multiple Attribute Group Decision Making under 2-Tuple Linguistic Neutrosophic Environment, Symmetry-Basel, 10 (2018) 15.

[41] C.E. Shannon, A mathematical theory of communication, Bell System

Technical Journal, 27 (1948) 379-423.

[42] A. Tversky, D. Kahneman, Prospect Theory: An Analysis of Decision under Risk, Econometrica, 47 (1979) 263-291.

[43] S. Huang, H. Chen, Research on quality evaluation of industry-education integration for rural vocational education in the perspective of rural revitalization with PL-MACONT method, Journal of Intelligent & Fuzzy Systems, 44 (2023) 9743-9755.

[44] P. Wang, J. Wang, G.W. Wei, J. Wu, C. Wei, Y. Wei, CODAS Method for Multiple Attribute Group Decision Making Under 2-Tuple Linguistic Neutrosophic Environment, Informatica, 31 (2020) 161-184.

[45] H. Liu, Performance evaluation of family business strategic transition based on the 2-tuple linguistic neutrosophic number multiple attribute group decision making, Journal of Intelligent & Fuzzy Systems, 44 (2022) 3271-3283.

[46] H.B. Zhang, An improved CLVA method for evaluating the endurance quality level of young male basketball players with 2-tuple linguistic neutrosophic information, Journal of Intelligent & Fuzzy Systems, 44 (2023) 9003-9014.

A Deep Learning Based Detection Method for Insulator Defects in High Voltage Transmission Lines

Wang Tingyu[1], Sun Xia[2*], Liu Jiaxing[3], Zhang Yue[4]
College of Intelligent Equipment, Shandong University of Science and Technology, Tai'an 27100, China[1, 2, 3]
College of Mechanical and Electrical Engineering, Zaozhuang University, Zaozhuang, 277160, China[4]

***Abstract*—The high-voltage transmission system is a key component of the power network, and the reliability of its insulators directly affects the safe operation of the system. Traditional insulator defect detection methods are reliant on manual inspection, which requires significant human resources and is prone to substantial subjectivity. To address this issue, this paper proposes an insulator defect recognition method based on the improved YOLOv5 algorithm. This method first collects images of insulator defects and then utilizes the YOLOv5 model for recognition training. To enhance multi-scale feature fusion capability, a bidirectional feature pyramid network (BiFPN) is introduced. During the training process, the SiUL function is used, and the SE attention mechanism has been integrated into the detection backbone network, which enhances the model's detection accuracy. Experimental results show that the model achieves a detection precision of 90.27%, a recall of 89.14%, and a mAP of 91.34% on the test set. To further enhance the model's practicality, a PyQt5-based user interface (GUI) for the inspection system is designed, enabling interactive functions such as image uploading, defect detection, and result display. In summary, the research presented in this paper provides efficient and accurate technical support for intelligent power inspection, offering a wide range of application prospects.**

***Keywords*—Insulators; insulator defect detection; improved YOLOv5; BiFPN network; PyQt5**

I. INTRODUCTION

High-voltage transmission lines are a key component of the power system, assuming the important task of transmitting electric energy, and their operational status is directly related to the stability and safety of the power system. As a key component of high-voltage transmission lines, insulators not only support and secure wires but also protect them from environmental erosion and mechanical damage. Therefore, the performance of insulators directly affects the insulation effect and service life of transmission lines. The traditional method for detecting insulator defects mainly relies on manual visual inspection, which has many problems [1]. First of all, the detection efficiency is low, manual visual inspection requires a lot of time and manpower, making it difficult to meet the needs of large-scale and rapid detection. Secondly, the misdetection rate is high; due to human factors, it is easy to miss or incorrectly detect defects, which poses potential risks to the safe operation of the power system.

With the continuous development of science and technology, deep learning-based image recognition technology shows great potential in the field of insulator defect detection. Deep learning technology can automatically extract defect features by learning from insulator images, facilitating efficient and accurate defect detection. Jia Yujin et al. (2023) proposed two lightweight enhancements to YOLOv5, combining the classical YOLOv5 with the advantages of the lightweight convolutional neural networks MobileNetV3 and GhostNet, respectively. Experimental results showed that the enhanced model reduced computational load by 49.4% while maintaining detection accuracy [2]. Ru Hongfang et al. (2023) proposed an improved YOLO x method for detecting insulator self-explosion defects, incorporating the CBAM attention mechanism into the backbone network and optimizing the IoU calculation of the loss function to EIoU, achieving a detection accuracy of 97.26% [3]. Satyajit et al. proposed an automated inspection system utilizing a six-rotor UAV and the YOLOv8n model, achieving efficient real-time monitoring by training the model with a dataset of 6020 insulator images and using image enhancement techniques to avoid overfitting. The YOLOv8n model achieved a mAP@50 of 99.4%, significantly enhancing the efficiency and accuracy of insulator detection [4]. Souza developed a Hybrid-YOLO model based on the ResNet-18 classifier, trained using 1593 grid inspection images, with a mAP of 0.99262 and an F1 score of 0.96216 for the multiclassification task, significantly improving the efficiency and accuracy of insulator detection [5]. Additionally, Yi et al. proposed the GC-YOLO model, which integrates the Ghost convolution module and CA attention mechanism in the backbone network and adds a small target detection head in the detection layer. Experimental results show that GC-YOLO achieves a recall of 89.7% and a mAP@0.5 of 94.2%, which are 7% and 6.5% higher than YOLOv5s, respectively [6]. These studies demonstrate that deep learning-based insulator defect detection techniques have significant advantages in improving detection efficiency and accuracy, showing promising application prospects.

Collectively, these studies had underscored the remarkable advantages of deep learning-based insulator defect detection technologies in enhancing detection efficiency and accuracy. However, the pursuit of further improvements in detection precision and speed remained a focal point of ongoing research. Against this backdrop, this paper proposed an insulator defect detection method for high-voltage transmission lines based on an enhanced YOLOv5 model. By incorporating the SE attention mechanism, the BiFPN module, and employing the SiLU loss function, along with the development of an insulator interaction system interface, this method aimed to achieve efficient and precise detection of insulator defects, thereby improving detection accuracy and reducing the false detection rate.

*Corresponding Author.

II. TESTS AND METHODS

A. Image Acquisition

In this study, the publicly available dataset of high-voltage transmission line insulators from Baidu AI Studio was used, along with insulator defect images obtained through web crawlers. Eventually, a total of 1033 insulator images were collected. Due to the limited number of insulator defect images in the original dataset, data augmentation techniques were applied to increase the dataset size and improve the model's generalization ability. After augmentation, the dataset contained 2066 images, some of which were shown in Fig. 1.

Fig. 1. Image of partial insulator defects.

B. YOLOv5 Model

The YOLOv5 model was the top-performing detection model in the YOLO family and consisted of four main modules: input, backbone network, neck, and head [7]. As shown in Fig. 2, the structure of the YOLOv5 model included techniques such as Mosaic data enhancement and adaptive anchor frame computation [8] in the input module. The backbone network employed Focus and CSPNet for feature extraction enhancement. The neck module combined different CSP modules and up-sampling techniques to obtain multi-scale contextual information. The head module was responsible for classification and regression tasks.

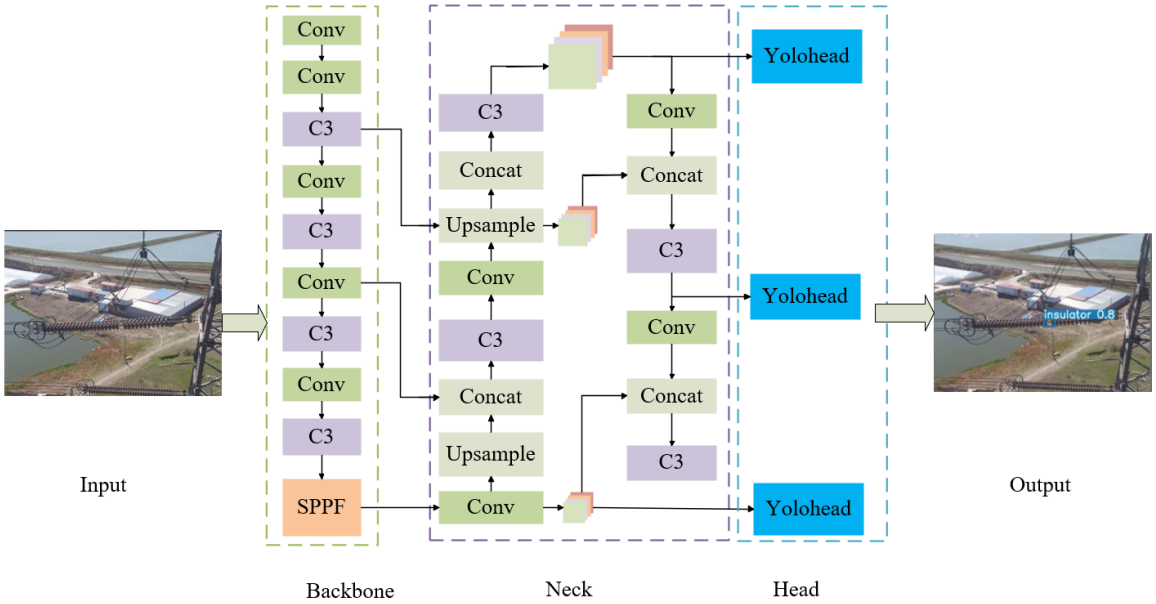


Fig. 2. YOLOv5 model structure.

C. YOLOv5 Improvements

1) SE attention mechanism: Squeeze-and-Excitation Networks (SE) [9] were implemented through two main steps: compression and excitation. As shown in Fig. 3, in the compression step, the SE module compressed the input feature map into a vector through a global average pooling operation, and then mapped it to a smaller vector through a fully connected layer. This process would be interpreted as summarizing and generalizing the overall information of the input features. In the excitation step, each element in this vector was scaled between 0 and 1 using a sigmoid function and multiplied with the original input feature map to obtain a weighted feature map [10]. This excitation operation would be understood as a recalibration of the local information of the input features. With the global averaged pooling and the excitation operations of the sigmoid function, the SE attention mechanism adapted to learn the importance of each channel, thus allowing the model to better understand the critical information in the input features and to focus on them more accurately in the output [11]. To improve the information interaction between each channel of the model and the utilization efficiency of feature information, this paper added the SE attention mechanism to the backbone network, and its specific position in the network was shown in Fig. 3.

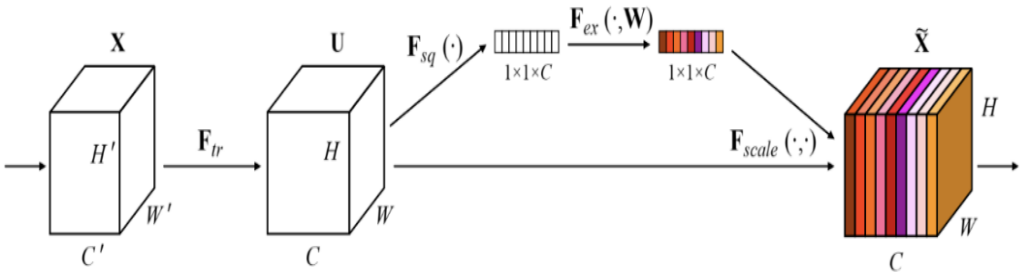


Fig. 3. Structure diagram of SE attention mechanism.

2) Bidirectional Feature Pyramid Network (BiFPN): BiFPN [12] achieved effective fusion of multi-scale features through both top-down and bottom-up pathways , with its structural diagram illustrated in Fig. 4. Its weighted feature fusion mechanism allowed the model to dynamically adjust the importance of different scale features based on task requirements, thereby enhancing feature representation accuracy. In the context of detecting insulator defects, which were typically small targets, a BiFPN module had been introduced into the model's neck section to replace the original feature pyramid network (FPN) [13]. This bi-directional feature fusion mechanism better integrated detailed information from low-level features with semantic information from high-level features, thereby improving the model's capability to detect defects across various scales.

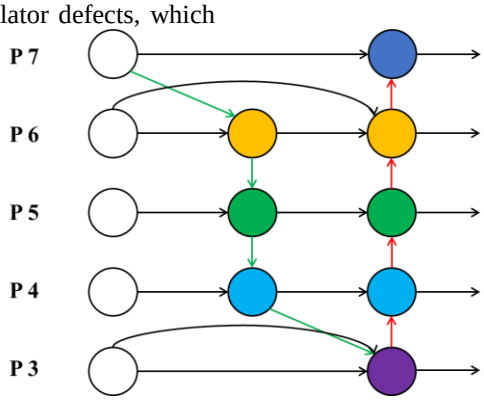


Fig. 4. Structure of bi-directional feature pyramid network.

3) SiLU activation function: In YOLOv5, the activation function introduced nonlinear factors, thereby enhancing the model's expressive power [14]. Activation functions playde a crucial role in neural networks by mapping neuron outputs to nonlinear intervals. This nonlinear mapping enabled neural networks to better adapt to complex patterns, thereby improving accuracy and performance. SiLU (Sigmoid-Weighted Linear Unit) acted as an implicit regularize with the following expression, suppressing the learning of numerous weights during training and enabling the model to focus more effectively on important features and patterns. This not only enhanced the model's generalization ability but also mitigates the risk of overfitting [15]. Additionally, due to SiLU's characteristics, the network's computational speed was improved. When the input x was large, SiLU's value was comparable to that of ReLU. Therefore, this paper adopted the SiLU activation function to enhance the overall performance of the model.

$$SiLU = x \times sigmoid(x) \quad (1)$$

$$sigmoid = \frac{1}{1+e^{-x}} \quad (2)$$

D. Interactive Interface Design for Detection Systems

In the development of an enhanced insulator defect detection system, designing an intuitive and user-friendly graphical user interface (GUI) was paramount. This section elaborated on the interactive interface design utilizing Python's PyQt5 library, as depicted in Fig. 5. The GUI aimed to streamline the interaction between users and the detection model, enhancing the system's usability and user experience.

The GUI integrated an image display area, an upload button, a detection button, and a feature for saving detection results in the backend. Users would effortlessly select an insulator image file for detection by clicking the upload button. Once loaded, the image was displayed in a designated area within the interface, facilitating user preview. Subsequently, upon clicking the detection button, the system automatically initiated the backend defect detection model, processing the image in real-time and instantly feeding back the detection results in graphical or textual form on the interface. This process not only visually showcased the detection effects but also significantly boosted detection efficiency.

To further augment the system's functionality and practicality, we had incorporated a logic for saving detection information in the backend after the detection process. Once

users confirmed the accuracy of the detection results, the information saving process was automatically triggered. The system then collected relevant detection details, including image paths, detection timestamps, defect types, locations, and severities, and leverages Python's file manipulation capabilities to save this information in a local JSON file. This step ensured the traceability and analyzability of detection results, facilitating subsequent data management and report generation.

Moreover, to guarantee a seamless user experience, meticulous optimization and testing of the GUI were conducted. By arranging interface elements logically, refining interaction logic, and incorporating error handling and user feedback mechanisms, we ensured the GUI's ease of use, stability, and reliability. Ultimately, a fully-featured, straightforward, and user-friendly GUI for the enhanced insulator defect detection system was successfully implemented, providing robust technical support for insulator inspection tasks in the power industry.

Fig. 5. Detection system interaction interface.

III. Results and Discussion

A. Test Environment

For this model training experiment, the environment was configured as follows: Python 3.8.16 served as the primary programming language with PyTorch 2.0.0 as the chosen deep learning framework. Additional software environments included Torch 2.0.0 and CUDA 11.8.0. The system ran on Windows 11, utilizing an Intel Core i7-13900HX CPU paired with an NVIDIA GeForce RTX 3060 graphics card and 16 GB of RAM. This setup ensure stable and efficient model training, guaranteeing smooth experimentation and accurate results [16].

B. Evaluative Indicators

In this study, multiple evaluative indicators were used to comprehensively assess the performance of the trained model. These indicators included precision, recall, and average precision, [17] which provided comprehensive information for evaluating different aspects of the model.

C. Optimization of Model Training and Validation Parameters

In the development of the insulator defect detection system, the optimization of parameters during the model training and validation stages represents a pivotal aspect ensuring model performance and generalization capability.

Firstly, regarding the learning rate configuration, it served as a crucial hyperparameter modulating the step size of model weight updates, significantly impacting training efficiency and stability. This system employed an initial learning rate of 0.001, coupled with a learning rate decay strategy. Specifically, as training progresses, the learning rate was automatically adjusted to 10% of its previous value every 10 epochs. This strategy aimed to balance rapid convergence during initial training phases with fine-tuning in later stages, thereby preventing the model from becoming trapped in local optima or overfitting.

Secondly, the selection of batch size needed a comprehensive consideration of hardware resource constraints and data processing efficiency. After thorough evaluation, a batch size of 32 was established in this system, ensuring efficient utilization of computational resources while maintaining the stability of the training process and mitigating noise in gradient estimations.

An initial number of 100 epochs was prescribed for training iterations, with ongoing monitoring of validation set performance to facilitate dynamic adjustments. If stagnation or a downward trend in validation set performance metrics (such as

loss or accuracy) was observed, it might indicate that the model is nearing optimality or beginning to overfit, necessitating timely termination of training.

In terms of optimizer selection, this system adopted the Adam optimizer, renowned for its adaptive learning rate adjustment capabilities, accelerating model convergence and minimizing computational resource consumption.

To further enhance model generalization, regularization techniques were integrated. Specifically, L2 regularization (weight decay) was employed, penalizing the squared sum of model weights to constrain model complexity. A weight decay coefficient of 0.0005 was set. Additionally, the Dropout strategy was implemented in select layers of the model, randomly discarding a proportion (0.3) of neuron outputs to bolster model robustness.

Through meticulous parameter optimization and effective training validation strategies, this system had successfully established a high-performance experimental environment for insulator defect detection models, laying a solid foundation for their stable performance and precise detection capabilities in practical applications.

D. Results and Analysis

1) Improved model validation analysis

a) Add SE attention mechanism analysis: To thoroughly evaluated the impact of incorporating SE attention mechanisms on the model's detection performance, this paper conducted a comparative analysis before and after their addition under identical hardware and software conditions. The specific results were presented in Table I, demonstrating that the inclusion of SE attention mechanisms notably enhances the model's accuracy in both feature extraction and classification. Specifically, precision increased by 1.3 percentage points, recall by 1.99 percentage points, and average precision by 1.79 percentage points.

TABLE I. COMPARISON OF TESTS BEFORE AND AFTER ADDING SE MODULE

Norm	Original model	After adding the SE module
Accuracy/%	88.72	90.02
Recall rate/%	86.36	88.35
Average precision/%	88.45	90.24

b) Adding Bidirectional Feature Pyramid Network (BiFPN) Analysis: To comprehensively assessed the impact of introducing a bidirectional feature pyramid network (BiFPN) on the model's detection performance, comparative experiments are conducted under identical hardware and software environments. The results, presented in Table II, demonstrate that the inclusion of BiFPN significantly enhances both the model's multi-scale feature fusion capability and its detection performance. Specifically, following the introduction of BiFPN, the precision rate increases by 1.06 percentage points, the recall rate by 2.7 percentage points, and the average precision by 1.92 percentage points.

TABLE II. COMPARISON OF TESTS BEFORE AND AFTER ADDING BIFPN

Norm	Original model	After adding BiFPN
Accuracy/%	88.72	89.78
Recall rate/%	86.36	89.06
Average precision/%	88.45	90.37

c) Analysis using the SiLU activation function: To evaluate the performance of the SiLU function, the activation function was tested before and after the enhancement of the YOLOv5 model on various datasets and models in this study. The results, presented in Table III, demonstrate that the SiLU activation function excels across all performance indicators. Specifically, the precision rate achieved 90.27%, the recall rate was 89.14%, and the average precision reached 90.34%.

TABLE III. COMPARISON OF TESTS BEFORE AND AFTER ADDING SE MODULE

Norm	Original model	SilU activation function
Accuracy/%	88.72	90.27
Recall rate/%	86.36	89.14
Average precision/%	88.45	90.34

Following the adoption of the SiLU function, the model's precision, recall, and average precision were further enhanced, underscoring SiLU's ability to optimize model parameters and improve generalization. The inclusion of the SE attention mechanism alongside the SiLU optimization function notably boost model performance. This synergy suggested that these dual enhancements effectively complement each other, resulting in improved overall model performance.

2) Comparative testing of different models: In this paper, several different models were tested against the improved YOLOv5 model. Specifically, YOLOv4, YOLOv5, and YOLOv8 were compared with the improved YOLOv5 model, and the results were shown in Table IV.

TABLE IV. TEST RESULTS OF DIFFERENT MODELS

Model	P/%	R/%	mAP/%
YOLOv4	86.35	85.26	86.88
YOLOv5	88.72	86.36	88.45
YOLOv8	90.12	88.95	90.22
This paper model	90.27	89.14	91.34

Based on the experimental results, the improved model in this study demonstrated superior performance compared to the YOLOv5, YOLOv4, and YOLOv8 models in terms of precision, recall, and average precision metrics. These findings underscored the effectiveness of the enhancement strategy proposed in this paper. Consequently, the improved model exceled across all performance metrics, affirming the efficacy of the enhancement strategy and providing robust support for future research and applications.

This paper compared and analyzed the image detection results of the YOLOv5 model on the test set before and after the

enhancement, as shown in Fig. 6. Fig. 6(a) illustrates the target detection results of the original YOLOv5 model, while Fig. 6(b) showd the target detection results of the improved YOLOv5 model proposed in this study. It was evident from the figures that the enhanced model achieved more accurate identification of insulator defects and significantly improves detection accuracy, particularly for smaller and more challenging defects. This demonstrated the effectiveness of the enhancement strategy in improving the performance of the YOLOv5 model, making it suitable for practical applications.

(a) (b)

Fig. 6. Comparison of models before and after improvement.

IV. Conclusion

This paper proposed an improved method for detecting insulator defects in high-voltage transmission lines, with a focus on intelligent identification. By incorporating a bidirectional feature pyramid network (BiFPN), the YOLOv5 model had been significantly enhanced in terms of multi-scale feature fusion. Additionally, the detection performance had been further optimized through the integration of an SE attention mechanism and the adoption of the SiLU activation function. Experimental validation had demonstrated that the improved model exhibited significant improvements in performance metrics such as precision, recall rate, and average precision, particularly in the areas of small target recognition and positional accuracy. Furthermore, the developed graphical user interface (GUI) inspection system enhances the model's usability, allowing users to easily upload images and perform defect detection operations.

However, it should be noted that this study also had potential limitations and constraints. For instance, the performance of the model might be affected by variations in lighting conditions and image quality. Future work could involve exploring more robust methods to address these challenges and further improving the model's detection capabilities.

References

[1] Huo Chao, Gu Xiaogang, Huang Lingqin, et al. Improved YOLO v4 model for insulator defect detection using aerial imagery[J]. Electronic Measurement Technology, 2023, 46 (09): 175-181. DOI:10.19651/j.cnki.emt.2211518.

[2] Jia Yujin, Zhang Zhencheng, Li Ximing, et al. Transmission line insulator defect detection based on lightweight YOLOv5 network [J]. Journal of Electric Power, 2024, 39 (01): 36-44. DOI:10.13357/j.dlxb.2024.005.

[3] Ru Hongfang, Zuo Chao, Han Long. Self-explosion defect detection method of insulators based on improved YOLOx[J]. Heilongjiang Electric Power,2023,45(02):124-128.

[4] Panigrahy S, Karmakar S. Real-time Condition Monitoring of Transmission Line Insulators Using the YOLO Object Detection Model with a UAV[J]. IEEE Transactions on Instrumentation and Measurement, 2024.

[5] Souza B J, Stefenon S F, Singh G, et al. Hybrid-YOLO for classification of insulators defects in transmission lines based on UAV[J]. International Journal of Electrical Power & Energy Systems, 2023, 148: 108982.

[6] Yi W, Ma S, Li R. Insulator and defect detection model based on improved YOLO-S[J]. IEEE Access, 2023.

[7] Gao Ang, Lu Chuanbing, Ren Longlong et al. Detection of apple blossom growth status based on improved YOLOX-s and validation analysis[J]. Chinese Journal of Agricultural Chemistry,2023,44(08):162-167.

[8] SiYu Chen, Zhangjie Fu. Fusion of efficient attention for multi-scale transmission line component detection [J]. Computer Engineering and Applications, 2024, 60 (01): 327-336.

[9] Hao Weiguang. Research on gas meter flow anomaly detection based on improved generative adversarial network [D]. Wuhan University of Science and Technology, 2023.

[10] Kunpeng Zhang. Research on the method of infected region segmentation of CT images of new coronary pneumonia based on deep learning [D]. Shanghai University, 2021.

[11] Wang Q,Wu B,Zhu P,et al.ECA-Net:Efficient channel attention for deep convolutional neural networks[C]//Procee-dings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2020:11534-11542.

[12] Zhu L, Deng Z, Hu X, et al. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 121-136.

[13] Deng C, Wang M, Liu L, et al. Extended feature pyramid network for small object detection[J]. IEEE Transactions on Multimedia, 2021, 24: 1968-1979.

[14] Zhang Xubo, Zhang Jiayu, Ben Zongyou et al. Research on the identification method of internal cracks in brown rice based on YOLO v5-RF[J/OL]. Journal of Nanjing Agricultural University,1-13.

[15] Dai Yunzhou, Lu Hong, Ji Chenyang. Grid worker defense detection based on improved YOLO v5[J]. Journal of Nanjing Engineering College (Natural Science Edition),2023,21(01):33-38

[16] Zou Huijun, Jiao Liangbao, Zhang Zhijian et al. Improved YOLO network for transmission line small target foreign object detection[J]. Journal of Nanjing Engineering College (Natural Science Edition),2022,20(03):7-14.

[17] Jubayer F, Soeb J A, Mojumder A N, et al. Detection of mold on the food surface using YOLOv5[J]. Current Research in Food Science, 2021, 4: 724-728.

A Review of Personalized Recommender System for Mental Health Interventions

Idayati Binti Mazlan[1], Noraswaliza Abdullah[2], Norashikin Ahmad[3], Siti Zaleha Harun[4]
Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia[1, 2, 3]
Advanced Computing Technologies Centre (C-ACT), Universiti Teknikal Malaysia Melaka (UTeM), Malaysia[1, 2, 3]
Faculty of Computing, Universiti Teknologi Malaysia (UTM), Malaysia[4]

***Abstract*—Personalized recommender systems for mental health are becoming indispensable instruments for providing individuals with individualized resources and therapeutic interventions. This study aims to explore the application of recommender systems within the mental health domain through a systematic literature review. The research is guided by three primary questions: 1) What is a recommender system, and what techniques are available within these systems? 2) What techniques and approaches are used explicitly in recommender systems for mental health applications? 3) What are the limitations and challenges in applying recommender systems in the mental health domain? The first step in answering these questions is to give a thorough introduction to recommender systems, covering all the different methods, including content-based filtering, collaborative filtering, knowledge-based filtering, and hybrid approaches. Next, examine the specific techniques and approaches employed in the mental health context, highlighting their unique requirements for adaptation, benefits, and limitations. Ultimately, the research highlights the key limitations and challenges, including data privacy concerns, the need for tailored recommendations, and the complexities of user engagement in mental health environments. By synthesizing current knowledge, this review provides valuable insights into the potential and constraints of recommender systems in supporting mental health, offering guidance for future research and development in this critical area.**

***Keywords*—Recommender system; collaborative filtering; content-based filtering; hybrid recommender system; mental health**

I. Introduction

Recommender systems, also called information filtering support, are used to predict or suggest an item, such as products, movies, content, and others, based on machine learning algorithms that the user might be interested in. It uses content-based filtering, collaborative filtering, and a hybrid recommender system to generate recommendations based on user preferences, behavior, and other relevant data. Today, recommendation systems have become the most popular tools for personalized support, and they have the potential to personalize mental health support further and improve treatment outcomes in the digital mental health landscape.

The area of mental health treatments has seen a significant change in recent years, with a focus on acknowledging the needs, preferences, and characteristics of those who are seeking help. One in three Malaysians have a mental disease, yet half of those diagnoses are erroneously raising concerns about mental health. Numerous mental health conditions are prevalent, such as anxiety, depression, bipolar disorder, schizophrenia, and more [1]. This shows that there is cause to be concerned and emphasizes the need for mental health support and education in Malaysia. While mental health interventions are essential for addressing people's well-being, the current field mainly uses long-standing, conventional methods, including clinical diagnosis and medication intervention [2]. Even while these traditional approaches are helpful, they have many limitations. The desire for personalization—where therapies can be customized to meet the unique needs of everyone—has been sparked by the realization that there is no one-size-fits-all approach to mental health care. This change reflects a broader understanding that individuals seeking help might benefit from more engaging, effective, and ultimately transformational mental health treatment.

The need for personalization in mental health interventions is becoming increasingly acknowledged, although a severe issue still exists. The way mental health treatment is now provided does not adequately allow therapies to be adapted to each patient's specific requirements. Although helpful, traditional methods might only fully utilize personalization's potential. This issue presents a significant difficulty since it might impair user engagement and satisfaction and, more importantly, the well-being of those who are seeking mental health help. Research on the effectiveness of recommender systems specifically designed for mental health therapies remains limited. The absence of robust empirical evidence makes it difficult to determine whether these approaches are beneficial and effective.

Thus, this study aims to conduct a systematic literature review on the application of recommender systems and their core techniques—including collaborative filtering, content-based filtering, knowledge-based filtering, and hybrid approaches—focusing specifically on their use in the mental health domain. It also seeks to compare these techniques, identifying their advances, challenges, and limitations associated with their implementation in this field.

II. Research Method

This paper offers an extensive systematic literature review that critically evaluates related works that pertain to the research topic. This dissertation's literature review mainly centers on diverse recommender systems that employ various recommendation techniques. These distinctions are organized and summarized in a tabulated format to enhance

comprehension. The primary objective of this systematic literature review is to evaluate the methods available for crafting a recommender system that is ideally tailored for personalized mental health interventions.

A literature review is essential to research that comprehensively analyses existing literature on a particular topic. In the case of recommender systems, literature reviews are conducted to identify the current state of research, research trends, limitations, and opportunities. It also provides a guideline for future research in recommender systems. Hence, this part is essential to understand the current recommender systems research comprehensively. Conducting a systematic literature review (SLR) involves several key stages. Fig. 1 shows the critical stages in conducting a systematic literature review for this study.

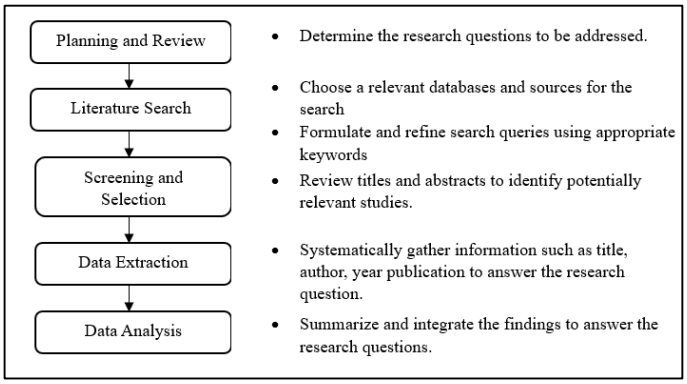


Fig. 1. Systematic literature review stage.

A. Planning and Review

This systematic review aims to understand the trends in recommender systems that can be effectively addressed, as well as the development and assessment of recommender systems. The research questions (RQ) aim to ascertain the limitations and difficulties associated with recommender systems and the evidence supporting their significance for individualized mental health. Below are the study questions that will be investigated during this comprehensive literature review:

- RQ1: What is a recommender system? What are the techniques available in the recommender system?
- RQ2: What techniques and approaches are used in recommender systems within the mental health domain?
- RQ3: What are the limitations and challenges in applying a recommender system for the mental health domain?

B. Literature Search

This section identifies relevant academic databases and sources well-regarded in personalized mental health recommender systems. Table I below shows an example of an academic database used to find relevant research in the field.

TABLE I. ACADEMIC DATABASE

Field	Academic Database
Recommender system	IEEE Xplore, ACM Digital Library, Scopus, and Google Scholar
Mental health interventions	PubMed.

The ultimate list of keywords was obtained after adding synonyms to make it longer. The search terms were defined using the research questions and the area's keywords as a guide. Table II presents examples of synonyms, and Table III presents a search query that was utilized for this study.

TABLE II. KEYWORD AND SYNONYM

Keyword	Synonym
Recommender	Suggestion, recommendation
System	Framework, mechanism, structure
Mental Health	Psychological health, emotional health, mental wellness
Intervention	Therapeutic support, support

TABLE III. SEARCH QUERY

Search ID	Search Term
S1	“Personalized recommender system” AND “mental health” AND “intervention”
S2	“Personalized recommendation for mental health” AND (“mental health intervention” OR “mental health therapeutic support”)
S3	“Customized mental health interventions” AND (“recommender system” OR “recommendation system”)
S4	(“Personalized recommendation” OR “adaptive recommendation”) AND (“mental health” OR “psychological health”) AND (“intervention” OR “support”)

C. Screening and Selection

In this activity, a few articles were selected to support the SLR. After identifying relevant papers, a quality assessment was conducted to avoid bias and internal and external validity. Incorporating the most recent research ensures that this SLR is current, comprehensive, and reflective of the state of the art in the application of recommender systems for mental health interventions. Therefore, only research published between 2019 and 2024 was chosen for this study. References were gathered, stored, and arranged using the desktop and web versions of the Mendeley software package.

About 612 results from the search were returned, comprising research articles, book chapters, proceeding papers, review articles, editorial materials, and early articles. After removing papers that were not relevant and adding relevant papers, 71) articles in total were used for this study. The flowchart below shows a detailed breakdown of the searched articles, as shown in Fig. 2.

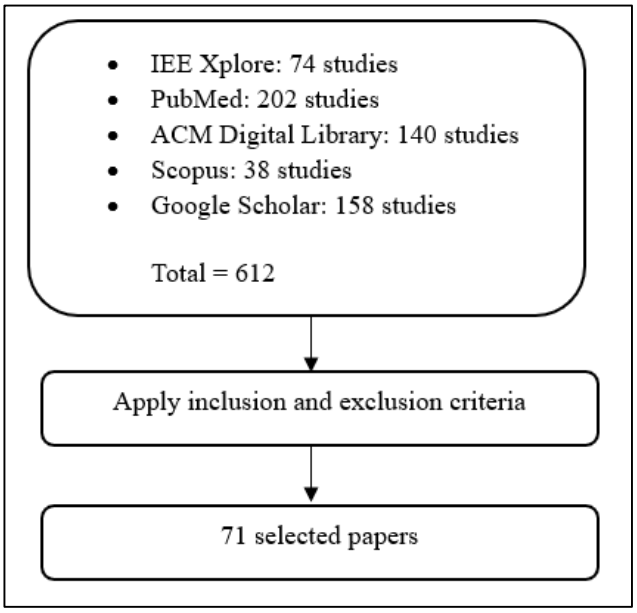


Fig. 2. Detailed breakdown of the searched articles

Explicit inclusion and exclusion criteria are necessary to guarantee that the studies are pertinent, excellent, and in line with the research goal. Table IV below shows the requirements for inclusion and exclusion.

TABLE IV. INCLUSION AND EXCLUSION

Inclusion	Exclusion
English article	Non-English article
Year published from 2019-2024	Outdated year publication
Related to the recommender system and mental health	Irrelevant to the studies
	Duplicate Literature

D. Data Extraction

Data extraction was carried out during the last phase of choosing primary research. Approaches, techniques, and application domains were among the crucial content data required to answer the research questions gathered, along with necessary metadata from the papers, such as author, title, and year of publication. Table V displays the framework for extracting data. First, a list of the extracted data is presented; next, any ambiguous data is clarified; finally, the research questions about the extracted information are explained in detail in the final column.

TABLE V. FRAMEWORK EXTRACTING DATA

Extract Data	Description	RQ
Author	Names of the authors of the study.	-
Title	Title of the study or paper	RQ1
Year of Publication	The year when the study was published	-
Journal/ Conference	Name of the journal or conference where the study was published.	-
Doi/ Link	Digital Object Identifier or URL for accessing the study	-
Application domain	Domain applied in the study	RQ2
Approach	Recommendation approach in the study	RQ3
Techniques	Technique applied in the study	RQ1

E. Data Analysis

For this review, 71 major research studies were dominated. From these studies, 23 journal articles, 26 conference papers, 4 workshop articles, 3 symposium articles, 4 handbook articles, and 11 others (website, review paper, etc.) were classified. The percentages of collated studies are presented in Fig. 3. In contrast, the total paper issued per year is presented in Fig. 4. This section presents the findings from a selection of studies that addressed each research question; these findings span several categories and include approaches, techniques, evaluation methodologies, and other crucial information.

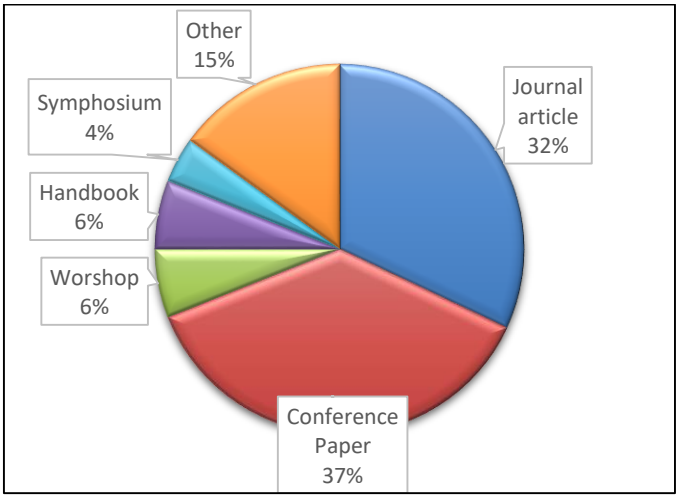


Fig. 3. Percentage of collated studies

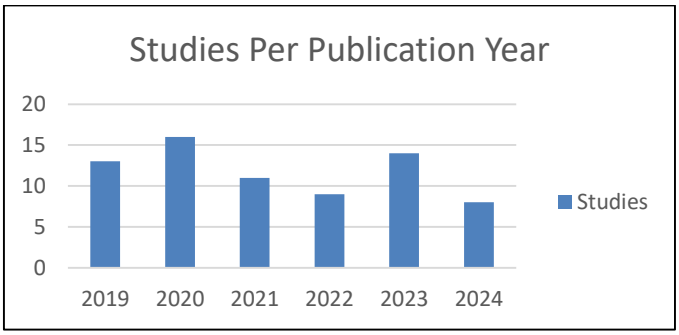


Fig. 4. Distribution of studies per publication year

III. REVIEW RESULT

A. Review Result Based on Research Question

1) Research question 1: What is a recommender system? What are the techniques available in the recommender system?

The rapid expansion of digital data and the growing number of internet users have led to the possibility of information overload, making it difficult for users to access relevant content on the web quickly. As a result, users often

struggle to find the information they need promptly, leading to frustration and decreased engagement with online services. This problem is particularly acute for users looking for highly specific or niche content, as they may have difficulty finding relevant items in the vast sea of available information. To address this challenge, recommender systems have been developed to help users find relevant and valuable content based on their past behaviors and preferences.

A recommender system is an information filtering system that predicts the preferences or interests of a user and recommends items that the user may like. It is widely used in e-commerce, social media, and entertainment industries to provide personalized recommendations to users [3]. Studies from [4] state that recommender systems have become more sophisticated in recent years with the development of deep learning algorithms and large datasets. Today, recommender systems are used in various applications, such as Netflix, which suggests movies based on the user's previous viewing history; YouTube recommends videos or tracks based on the user's search history; and Amazon suggests additional products based on the user's purchase history. Similarly, Facebook recommends that people befriended based on the user's social network. Traditional recommender systems use different approaches to make recommendations to users. These approaches include content-based filtering, collaborative filtering, hybrid methods, and other techniques that can be classified into more categories, such as knowledge-based filtering [4].

The area of mental health treatments has seen a significant change in recent years, focusing on acknowledging the particular needs, preferences, and characteristics of those seeking help. One in three Malaysians have a mental disease, yet half of those diagnoses are erroneously raising concerns about mental health. Numerous mental health conditions are prevalent, such as anxiety, depression, bipolar disorder, schizophrenia, and more [5]. This shows that there is cause to be concerned and emphasizes the need for mental health support and education in Malaysia. While mental health interventions are essential for addressing people's well-being, the current field mainly uses long-standing, conventional methods, including clinical diagnosis and medication intervention [2]. Even while these traditional approaches are helpful, they have a lot of limitations. The desire for personalization—where therapies can be customized to meet the unique needs of each individual—has been sparked by the realization that there is no one-size-fits-all approach to mental health care. This change reflects a broader understanding that individuals seeking help might benefit from more engaging, effective, and ultimately transformational mental health treatment.

Thus, by utilizing recommender systems, which have become valuable instruments in the more significant healthcare industry, data-driven solutions that can improve personalized treatment are provided. These systems employ user data and algorithms to offer suggestions to enhance user experiences and healthcare service delivery. While this usage offers convenience, some disadvantages need to be addressed. Therefore, this study uses a hybrid recommendation system that combines several recommendation techniques, such as content-based filtering, collaborative filtering, knowledge-based filtering, and hybrid approach, which may overcome the problems and provide more accurate and diverse recommendations.

a) Collaborative filtering approach: Recommender systems use a technique called Collaborative Filtering (CF) to forecast a user's interests based on the tastes of users who are similar to the said user [6]. It is predicated on the notion that individuals who share a joint assessment of a given item will likely do so in the future. User-based and item-based CF are the two primary categories of CF. While item-based CF suggests products comparable to those a user has previously liked, user-based CF suggests items to a user that similar users have previously liked [7].

As to [8], user-based CF is a method that forecasts the items a user would find appealing by analyzing the ratings that other users with comparable tastes have assigned to that item. This kind of recommendation system bases its product recommendations on user similarities. On e-commerce platforms, this technique is frequently used to make product recommendations to consumers based on their browsing and past purchases. To suggest movies, TV series, and songs to customers based on their viewing or listening history is also utilized by streaming services like Netflix and Spotify [9].

Item-based CF is a subset that makes recommendations to consumers based on how two comparable items are determined by utilizing user ratings. [10] states that Amazon utilized and invented it in 1998, which has been essential to the company's success. Item-based CF examines the relationship between the two items and combines the user's rated and bought items with related goods to generate a suggestion list.

Hence, CF can be classified as a powerful tool for personalizing users' experiences and increasing customer engagement and satisfaction. Table VI below shows several advantages and disadvantages of CF recommendation approaches for recommending items.

TABLE VI. ADVANTAGES AND DISADVANTAGES OF COLLABORATIVE FILTERING APPROACH

Advantage	Disadvantage
Personalization: CF provides users with personalized recommendations based on their past behavior, preferences, and activities [11].	Cold start problem: CF has a cold start problem, in which it isn't easy to make recommendations for new users or items without ratings [12].
Diversity: CF can recommend items that are not popular, which can increase the diversity of recommendations and expose users to new items [12].	Sparsity: CF can suffer from sparsity, where there are not enough ratings or data points to make accurate recommendations [12].
Scalability: CF can handle large datasets and many users, making it scalable for e-commerce platforms and other applications [11].	Over-specialization: CF can lead to over-specialization, where users are recommended items that are too similar to their past behavior, limiting their exposure to new items [12].
Adaptability: CF can adapt to user interest and preference changes over time, making it a dynamic and flexible recommendation system [11].	Popularity bias: CF can suffer from popularity bias, where popular items are recommended more frequently, leading to a need for more diversity in recommendations [11].

b) Content-based filtering approach: In recommender systems, content-based filtering (CBF) is a technique that uses the user's attributes to forecast and suggest new but comparable things [13]. The more a consumer participates, the more accurate future suggestions are generated by the CBF algorithm, which retains past user data, including clicks, reviews, and favorites, to establish a user profile. The author in [14] claims that this method does not use the data of other users when making suggestions to a single user; instead, it attempts to infer a user's features or behavior based on the item's features. Each item's unique traits must reflect its fundamental characteristics when using CBF. For example, the recommender system requires specific movie attributes, including actors and actresses, directors, the year of release, and genre, to discriminate between movies [15].

According to a [18] publication, the CBF model finds similarities between things to make suggestions based on specific criteria. These systems generate data profiles by utilizing description information, which could comprise attributes of objects or individuals. The users who previously liked, purchased, viewed, or listened to products are then suggested to them based on the developed profiles. The premise is that people may enjoy similar goods in the future if they have previously liked specific items, which is crucial to content-based filtering.

Two common approaches are typically employed in content-based filtering: Cosine distance and classification [13] [17]. In the cosine distance method, the cosine distance between the user and item vectors determines preference. For instance, the action movie vector has a positive value for a user who enjoys watching it, but the horror movie vector has a negative value for that same person. The cosine distance is used to find the similarity between the user and item vectors, and the items with the highest similarity score are recommended for use [13] [18]. The classification approach method uses a classification algorithm to predict whether a user will like an item or not based on the item's features; for example, a decision tree algorithm can be used to predict whether a user will like a movie based on its genre, director, actors, and other features. The algorithm learns from the user's past preferences and creates a model that can predict the user's future preferences [13] [18].

In summary, CBF is a method for recommender systems that uses machine learning algorithms to forecast and suggest new but comparable items based on those attributes to the user. The more a consumer participates, the more accurate future suggestions are generated by the CBF algorithm, which retains past user data, including clicks, reviews, and favorites, to establish a user profile. Rather than focusing on how an object interacts with the user, CBF approaches need additional information about its features. Table VII below shows several advantages and disadvantages of CBF recommendation approaches for recommending items.

TABLE VII. ADVANTAGES AND DISADVANTAGES OF CONTENT-BASED FILTERING

Advantage	Disadvantage
No-item cold-start problem: Content-based filtering can recommend items to new users who have yet to have any interaction [19].	Limited capacity to build upon users' current interests: The model can only provide recommendations based on the user's current interests. It is not able to build upon those interests [19].
Scalability: Since the recommendations are unique to this person, the model does not require any information about other users, simplifying scaling to many users [19].	Needs extensive domain expertise: Because the item feature representation is partially hand-engineered, this technique calls for a great deal of subject knowledge [19].
Recommendations for niche products: Based on the user's preferences, the model can identify products that only a tiny percentage of users are interested in [19].	Limited quality consideration: Content-based filtering systems mostly need to consider the quality of the items in the recommendation process [20]

c) Knowledge-based filtering approach: The knowledge-based filtering (KBF) approach is a recommendation system that does not rely on user data or ratings but instead uses domain knowledge to make recommendations. KBF systems use a set of rules or knowledge to recommend items to users based on their preferences [21][23].

According to [22], KBF systems are often used in domains where domain knowledge is readily available, such as the medical field or in recommending books. However, the quality and completeness of their domain knowledge need to be improved.

In summary, the knowledge-based filtering approach is a recommendation system that uses domain knowledge to make recommendations. It does not rely on user data or ratings, can explain its recommendations, and can handle new items. However, it is limited by the quality and completeness of its domain knowledge; it can only recommend items that match the user's known preferences and acquiring domain knowledge can take time and effort. Table VIII below shows several advantages and disadvantages of KBF recommendation approaches for recommending items.

TABLE VIII. ADVANTAGES AND DISADVANTAGES OF KNOWLEDGE-BASED FILTERING

Advantage	Disadvantage
No need for user data: Knowledge-based filtering systems do not require user data or ratings, making them useful when user data is unavailable [21][2].	Limited to domain knowledge: Knowledge-based filtering systems are limited by the quality and completeness of their domain knowledge [22].
Able to offer clarifications: Knowledge-based filtering systems can make their recommendations more visible and understandable by providing justifications for them [22].	Limited to known preferences: Knowledge-based filtering systems can only recommend items that match the user's known preferences and cannot expand on the user's interests [23].
Can handle new items: Knowledge-based filtering systems can handle new items that have yet to be rated by users [23].	Difficulty in acquiring domain knowledge: Acquiring domain knowledge can be challenging and time-consuming [22].

d) Hybrid recommender system approach: Hybrid recommender systems integrate the complementary qualities of two or more recommendation algorithms differently to benefit from the complementary qualities of two or more recommendation algorithms. Combining two or more recommendation algorithms, hybrid recommendation systems seek to give users more accurate, varied, and well-balanced recommendations. [24][25].

Developing hybrid recommender systems by fusing various methods, like content-based and collaborative filtering, is possible. While the content-based filtering component considers objects' particular characteristics and properties, the collaborative filtering component gathers the general public's knowledge [25][26]. According to an article by [24], there are various kinds of hybrid recommender systems, such as feature combination hybrid, weighted hybrid, switch hybrid, and mixed hybrid.

In summary, hybrid recommender systems leverage the complementary advantages of two or more recommendation algorithms by combining them in diverse ways. They can offer better coverage of the item space, overcome the shortcomings of individual methodologies, and deliver more precise recommendations. Nevertheless, choosing the best mix of approaches can take time and effort, and they might be more challenging to develop and maintain than single-technique systems. They also require more data. Different techniques, such as collaborative and content-based filtering, can be combined to create hybrid recommender systems. Table IX below shows several advantages and disadvantages of the hybrid recommender system approach in recommending items.

TABLE IX. ADVANTAGES AND DISADVANTAGES OF HYBRID RECOMMENDER SYSTEM

Advantage	Disadvantage
Overcoming limitations of individual techniques: Hybrid recommender systems can combine individual techniques to overcome their limitations [25].	Complexity: Hybrid recommender systems can be more complex to implement and maintain than single-technique systems [25].
Increased accuracy: Hybrid recommender systems can produce more accurate recommendations by integrating the advantages of several strategies [27].	Data requirements: Hybrid recommender systems require more data than single-technique systems [25].
Better coverage: Hybrid recommender systems can combine different techniques to provide better coverage of the item space [24].	Difficulty in selecting the right combination of techniques: Selecting the right combination of techniques can be challenging, and the hybrid system's performance depends on the quality of the individual methods [24].

2) Research question 2: What techniques and approaches are used in recommender systems within the mental health domain?

Recommender systems can improve the user experience of mental health apps by personalizing the intervention, making it more applicable to the needs of the individual user and, thus, more engaging [28] [29]. Through personalized recommendations, these systems can effectively customize the content, resources, and interventions offered by mental health apps, creating a more tailored and relevant experience for each user. According to [29], recommender systems can filter content and provide tailored mental health suggestions by leveraging individual usage data. Through meticulous analysis of user behavior, preferences, and interactions within mental health apps, these systems can discern patterns and preferences unique to each user. This insightful data analysis enables recommender systems to curate personalized content, ensuring that mental health suggestions are finely tuned to match the specific needs and interests of the individual.

Other than that, the recommender system for mental health personalization represents a transformative approach in mental health care, enabling the adaptation of treatments to the unique needs of individuals and thereby enhancing the overall care for depressive symptoms and beyond [30]. By customizing therapeutic strategies based on a person's specific circumstances, preferences, and responses, these interventions can optimize the effectiveness of treatment. According to [31], content recommendation systems are pivotal in scaling and complementing digital mental health care by offering personalized content and self-care recommendations. These systems leverage advanced algorithms to analyze user data, preferences, and behaviors, tailoring the delivery of relevant resources to individuals seeking mental health support.

Therefore, recommender systems can transform the mental health care field by personalizing interventions and better tailoring them to specific users' needs. They can be applied in various ways in mental health apps to ascertain what would be pertinent to the user. They use algorithms to forecast material or information relevant to the user.

a) Collaborative filtering for mental health: A CF Recommender System for Mental Health technology uses algorithms to provide personalized recommendations for mental health treatments or interventions. This system is designed to enhance the personalization of mental health care by recommending treatments or interventions based on the similarities between patients and their responses to different treatments. According to [22], CF algorithms, such as matrix factorization and k-nearest neighbors, have been used to improve the personalization of digital mental health therapy. These algorithms have shown a 6.5-8.3% improvement in treatment personalization.

The system uses patient ratings of treatment efficacy and demographic data to recommend treatments to new patients. Based on shared demographic and behavioral traits, it groups users into clusters. The model may suggest a treatment for a patient who is entirely new to the model or has taken depressive therapies in the past by comparing and classifying users based on specified demographic similarities and comparing treatments based on efficacy rating similarities across users [32].

Furthermore, [33] points out that CF can suggest places and activities that, according to users' prior experiences, may have a greater chance of having a favourable impact. With this, users may find tiny moments of thankfulness and joy in their daily lives. Recommended systems for digital mental health apps have some ethical concerns, especially for those that employ artificial intelligence and personal data. These include concerns

about privacy, personalization trade-offs, explainability, and the calibre of recommendations [22].

In summary, by tailoring interventions to the specific needs of each user, CF Recommender Systems for Mental Health has the potential to transform the way mental health care is provided completely. However, these systems' efficacy depends on the quantity and calibre of data available, and their application necessitates careful consideration of moral dilemmas.

b) Content-based filtering for mental health: A CBF Recommender System for Mental Health is a technology that uses algorithms to provide personalized recommendations for mental health treatments or interventions based on individual usage data. This system is designed to enhance the personalization of mental health care by recommending treatments or interventions based on the individual's past behavior and preferences. The author in [29] claim that CBF approaches have great potential for offering individualized mental health advice. Based on a user's personal information and past usage statistics, these strategies employ information filtering algorithms to provide suggestions or recommendations for appropriate material.

Therefore, by tailoring interventions and making them more relevant to the needs of specific users, CBF Recommender Systems for Mental Health have the potential to transform the way that mental health care is provided completely. However, the quality and availability of data determine how effective these systems are and using them requires careful ethical considerations [29].

Unfortunately, minimal research has been conducted in this area, possibly due to several reasons, including the sensitive nature of mental health data and the recent emergence of technology-driven solutions. It is possible that this area of study is still emerging or that research in this specific niche needs to be expanded.

c) Hybrid recommender system for mental health. A hybrid recommender system for mental health combines multiple recommendation approaches to provide individuals with personalized mental health support and resources. Research by [33] about a hybrid recommender system for mental illness detection in social media using deep learning techniques mentions that a hybrid recommender system is a sophisticated tool that combines various algorithms to analyze social media data and identify potential signs of mental illness. This system is primarily designed to optimize the detection of mental health issues, particularly depression, by analyzing patterns in social media behavior and content.

However, while hybrid recommender systems have demonstrated their potential in various domains, there is a notable lack of research specifically focused on their application in mental health. Technology plays a crucial role in mental health support, so developing hybrid recommender systems can contribute to more effective, accessible, and user-centric mental health care.

3) Research Question 3: What are the limitations and challenges in applying a recommender system for the mental health domain?

The sensitive nature of mental health data and the complexity of individual needs make the application of recommender systems in the mental health domain complex and fraught with limitations. Technological, data science, ethics, and mental health specialists must collaborate to address these issues multidisciplinary. Table X below shows the constraints and challenges in several studies on the recommender system approach using the CBF technique, CF technique, and hybrid recommender system in the mental health domain.

TABLE X. SUMMARIZATION OF RECOMMENDER SYSTEM FOR MENTAL HEALTH DOMAIN

Recommender System Approach	Studies	Limitation and challenge
CBF Technique	[29]	The lack of explainability in recommender systems used in DMH apps raises ethical concerns.
CF Technique	[27]	The analysis does not consider the cold-start scenario, in which the algorithms predict for users something they have never seen before, which is a well-known challenge for recommender systems.
	[31]	● User cold-start problem: User-user collaborative filtering relies on finding similarities in item ratings between users to make recommendations. However, a cold-start user with no item ratings cannot be given a recommendation. ● The sparsity of data: If users have rated only a few items, similarity measures between users become unreliable, affecting the accuracy of recommendations.
	[32]	There are challenges in developing engaging and personalized digital interventions, the potential of recommender systems to address these challenges, and the need for further research and empirical evaluation.
Hybrid Recommender System	[33]	Requires large labelled social media datasets, which can be difficult to obtain

IV. DISCUSSION

Specific gaps and limitations have been evident in the recommender systems literature, pointing to areas where further exploration and development are warranted. One of the primary challenges is the availability and quality of the data. Recommendation algorithms may not be as successful when dealing with sparse or noisy mental health data, such as user preferences, behavior patterns, and treatment results. To address this issue, data are collected through an online platform, which offers users the most accessible means to contribute and receive recommended interventions from the system. Experts in the field of mental health provide these intervention data. Using an online platform to collect data and incorporating the expertise of mental health professionals could

help alleviate some of the difficulties associated with limited or unreliable mental health data. However, it is essential to recognize that these approaches may not eliminate all problems. Nevertheless, they have the potential to significantly improve the quality, relevance, and effectiveness of the system's proposed mental health therapies.

Furthermore, it is crucial to acknowledge the limited research on recommender systems designed for mental health treatments. Prior research in this field often utilizes several recommendation strategies, demonstrating a wide array of methodologies. Yet, these methods have their inherent restrictions and difficulties. While hybrid recommender systems have the potential to tackle some obstacles, there is still a lack of research on their use in mental health settings. Although hybrid recommender systems have the potential to be effective in the field of mental health intervention, more research is required to evaluate their practicality, effectiveness, and implementation. The potential for synergies in hybrid recommender systems still needs to be explored. Focusing only on specific methods can make it harder to understand hybrid approaches, which could mean missing chances to improve the quality of recommendations.

Moreover, consistency in evaluating metrics across studies presents challenges in comparing and benchmarking recommender system performance. Standardized metrics and benchmarks are lacking, making it difficult to establish a common framework for evaluating recommendation quality. Addressing this requires the adoption of standardized metrics and benchmarks for more equitable comparisons.

A common challenge in recommender systems is the "cold start" problem, where recommender systems struggle to make practical recommendations for new users or items with limited data. Many studies need to adequately address this issue, which can lead to suboptimal recommendations for newcomers. Solving this problem is essential to ensure valuable recommendations for all users. In this case, combining a few techniques, which are CF, CBF, and KBF, may leverage the strengths of each approach to overcome the limitations of individual methods and handle the cold start problem more effectively.

In conclusion, it is essential to identify these holes and problems in the literature on recommender systems and take action so that the field can continue to grow and recommendation technologies improve. Future research on these areas will contribute to advancing recommender systems and their ability to serve diverse user needs and preferences.

V. Conclusion

This paper comprehensively explores recommender systems, focusing on enhancing personalization. Recommender systems are central in personalizing recommendations across various domains, improving user experiences, and guiding choices. Traditional approaches like CF and CBF are commonly used but have limitations, including the cold start problem and difficulty capturing diverse user preferences. There are challenges in delivering personalized care, including privacy concerns, the need for real-time recommendations, and limitations in cross-domain recommendations. Hence, hybrid recommender systems, which combine various recommendation techniques, hold promise for overcoming the limitations of traditional methods.

Evaluating recommender systems is essential, and standardized metrics and benchmarks are necessary for fair assessments. User and intervention data should be integrated into the evaluation of the recommendation system. Additionally, interpretability and explainability of recommendations are critical to building user trust and engagement. Privacy and ethical considerations, including data misuse and user profiling, should be considered to ensure responsible data use.

To wrap up, this review provides valuable insights into the world of recommender systems, their challenges, and the potential for improvement. The review emphasizes the importance of addressing gaps and limitations to create more effective and ethical recommendation systems that meet users' diverse needs and preferences.

VI. Future Work

Insights from the literature review suggest several promising directions for future research in recommender systems for mental health therapies. These directions are essential for overcoming current limitations and advancing the development of practical, user-friendly, and ethically sound recommendation technologies in mental health care.

Future work should prioritize addressing key challenges, including improving recommendation accuracy, mitigating cold start and data sparsity problems, and upholding ethical standards in system design and deployment. Enhancing the personalization of recommendations by integrating contextual factors such as user preferences, emotional states, and mental health history could significantly improve system performance. Moreover, innovative approaches to data utilization, such as incorporating external data sources and leveraging hybrid models, could help alleviate data-related limitations.

Research should also focus on evaluating the real-world efficacy of these systems through longitudinal studies and clinical trials to better understand their impact on mental health outcomes. Additionally, expanding the applicability of recommender systems to diverse populations is crucial for ensuring inclusivity and addressing cultural, demographic, and individual differences in mental health needs.

Exploring advanced techniques, such as deep learning, reinforcement learning, and multimodal data integration, offers further opportunities to push the boundaries of current methods. By investigating these areas, future research can significantly contribute to developing more accurate, portable recommender systems that align with ethical considerations and user needs.

Acknowledgment

The authors are grateful to the Universiti Teknikal Malaysia Melaka (UTeM) for supporting this research.

References

[1] Mental health in Malaysia: Overcoming stigma and benefits of improving access to care. Mental Health Malaysia: Overcoming Stigma |

Prudential Malaysia. (n.d.). https://www.prudential.com.my/en/we-do-pulse/all-stories/mental-health-malaysia/

[2] Siddiqui, S., Morris, A., Ikeda, D. J., Balsari, S., Blanke, L., Pearsall, M., Rodriguez, R., Saxena, S., Miller, B. F., Patel, V., & Naslund, J. A. (2022). Scaling up community-delivered mental health support and care: A landscape analysis. Frontiers in public health, 10, 992222. https://doi.org/10.3389/fpubh.2022.992222

[3] Kim, J., & Lee, J. (2021). The Impact of Accuracy and Diversity on Customer Satisfaction with Recommender Systems. Journal of Information Science Theory and Practice, 9(2), 1-14.

[4] Sharma, S., Gupta, K., & Gupta, D. (2021). Recommender System: A bibliometric analysis. In IOP Conference Series: Materials Science and Engineering (Vol. 1022, No. 1, p. 012057). IOP Publishing.

[5] Mahtar, S.N.A.M., Masrom, S., Omar, N., Khairudin, N., Rahim, S.K.N.A. and Rizman, Z.I. (2017). Trust-aware recommender system with distrust of different views of trusted users. Journal of Fundamental and Applied Sciences, 9(5S), pp.168-182.

[6] 7Gupta, S. (2019, April 1). Collaborative filtering and embeddings - part 1. Medium. https://towardsdatascience.com/collaborative-filtering-and-embeddings-part-1-63b00b9739ce

[7] [8] Ajitsaria, A. (2022). Build a recommendation engine with collaborative filtering, Real Python. Available at: https://realpython.com/build-recommendation-engine-collaborative-filtering/ (Accessed: 23 June 2023).

[8] Gupta, R. (2023, January 22). User-based collaborative filtering. GeeksforGeeks. https://www.geeksforgeeks.org/user-based-collaborative-filtering/

[9] Konduforov, A. (2021, July 27). Recommender systems: Behind the scenes of machine learning-based personalization. AltexSoft. https://www.altexsoft.com/blog/recommender-system-personalization/

[10] Qutbuddin, M. (2023, September 3). A comprehensive guide on Item Based Recommendation Systems. Medium. https://towardsdatascience.com/comprehensive-guide-on-item-based-recommendation-systems-d67e40e2b75d

[11] Holewa, K. (2020, June 29). We know what you like! Perks of recommendation systems in business. Medium. https://medium.com/swlh/we-know-what-you-like-perks-of-recommendation-systems-in-business-5f227bb6d09

[12] Skovhøj, F. Z. (2022, February 9). Using collaborative filtering in e-commerce: Advantages & disadvantage. Clerk.io. https://www.clerk.io/blog/collaborative-filtering

[13] Turing Enterprises Inc. (2022, August 5). A guide to content-based filtering in Recommender Systems. A Guide to Content-based Filtering in Recommender Systems. https://www.turing.com/kb/content-based-filtering-in-recommender-systems

[14] Roy, A. (2020, July 31). Introduction to Recommender systems- 1: Content-based filtering and collaborative filtering. Medium. https://towardsdatascience.com/introduction-to-recommender-systems-1-971bd274f421

[15] Castells, P. and Jannach, D. (2023). Recommender Systems: A Primer. arXiv preprint arXiv:2302.02579.

[16] AltexSoft. (2021, July 27). Recommender systems: Behind the scenes of machine learning-based personalization. AltexSoft. https://www.altexsoft.com/blog/recommender-system-personalization/

[17] The Upwork Team. (2021, April 6). What content-based filtering is & why you should use it | upwork. https://www.upwork.com/resources/what-is-content-based-filtering

[18] Leban, J. (2020, May 21). Essentials of recommendation engines: Content-based and collaborative filtering. Medium. https://towardsdatascience.com/essentials-of-recommendation-engines-content-based-and-collaborative-filtering-31521c964922

[19] Google. (n.d.). Collaborative Filtering Advantages & disadvantages | machine learning | Google for developers. Google. https://developers.google.com/machine-learning/recommendation/collaborative/summary

[20] Lops, P., Jannach, D., Musto, C., Bogers, T., & Koolen, M. (2019). Trends in content-based recommendation: Preface to the special issue on Recommender systems based on detailed item descriptions. User Modeling and User-Adapted Interaction, 29, 239-249.

[21] Nießner, J., & Ludwig, T. (2021). Design of a knowledge-based recommender system for recipes from an end-user perspective. Mensch Und Computer 2021. https://doi.org/10.1145/3473856.3473888

[22] Markchom, T., Liang, H., & Ferryman, J. (2023). Scalable and explainable visually-aware Recommender Systems. Knowledge-Based Systems, 263, 110258. https://doi.org/10.1016/j.knosys.2023.110258

[23] Chicaiza, J., & Valdiviezo-Diaz, P. (2021). A comprehensive survey of knowledge graph-based recommender systems: Technologies, development, and contributions. Information, 12(6), 232. https://doi.org/10.3390/info12060232

[24] Çano, E. and Morisio, M. (2017). Hybrid recommender systems: A systematic literature review. Intelligent Data Analysis, 21(6), pp.1487-1524.

[25] Kharwal, A. (2023, June 5). Hybrid recommendation system using Python: Aman Kharwal. thecleverprogrammer. https://thecleverprogrammer.com/2023/06/05/hybrid-recommendation-system-using-python/

[26] Lineberry, A. (2018, September 11). Creating a hybrid content-collaborative movie recommender using Deep Learning. Medium. https://towardsdatascience.com/creating-a-hybrid-content-collaborative-movie-recommender-using-deep-learning-cc8b431618af

[27] Parks, J. (2023, April 27). How to Build a Hybrid Recommender System. How to build a hybrid recommender system. https://agilethought.com/blogs/how-to-build-hybrid-recommender-system/

[28] Lewis, R. et al. (2022) 'Can a recommender system support treatment personalisation in Digital Mental Health therapy? A quantitative feasibility assessment using data from a behavioural activation therapy app,' CHI Conference on Human Factors in Computing Systems Extended Abstracts [Preprint]. doi:10.1145/3491101.3519840.

[29] Valentine, L., D'Alfonso, S. and Lederman, R. (2022). Recommender systems for mental health apps: advantages and ethical challenges. AI & Society, pp.1-12.

[30] Hornstein, S., Zantvoort, K., Lueken, U., Funk, B., & Hilbert, K. (2023). Personalization strategies in digital mental health interventions: A systematic review and Conceptual Framework for depressive symptoms. Frontiers in Digital Health, 5. https://doi.org/10.3389/fdgth.2023.1170002

[31] Chaturvedi, A., Aylward, B., Shah, S., Graziani, G., Zhang, J., Manuel, B., Telewa, E., Froelich, S., Baruwa, O., Kulkarni, P. P., Ξ, W., & Kunkle, S. (2023). Content Recommendation Systems in Web-Based Mental Health Care: Real-world Application and Formative Evaluation. JMIR formative research, 7, e38831. https://doi.org/10.2196/38831

[32] Orlando, A., Venuti, K., & Tesfalul, M. (n.d). Collaborative Filtering Recommender System for Treatment of Depression.

[33] Siriaraya, P., Suzuki, K., & Nakajima, S. (2019, September). Utilizing Collaborative Filtering to Recommend Opportunities for Positive Affect in daily life. In HealthRecSys@ RecSys (pp. 2-3).

[34] Ahmad, S. S., Rani, R., Wattar, I., Sharma, M., Sharma, S., Nair, R., & Tiwari, B. (2023). Hybrid recommender system for mental illness detection in social media using deep learning techniques. Computational Intelligence and Neuroscience, 2023(1), 8110588.

Intelligent Service Book Sorting in University Libraries Based on Linear Discriminant Analysis Method

Changjun Wang*, Fengxia You, Yu Wang
Library, University of Sanya, Sanya, 572022, China

***Abstract*—The demand for intelligent services in university libraries is constantly increasing, especially in the intelligent book sorting work. The research aims to explore an intelligent classification method for university library books based on linear discriminant analysis. It is used to reduce the dimensionality of feature multidimensional data. A membership model for different categories of books is established to achieve classification. The results showed that when the training set data was reduced to two-dimensional, the feature extraction accuracy of the classification algorithm reached 64.02%, which was significantly higher than 52.48% of one-dimensional data. In addition, the membership calculation accuracy of axiomatic fuzzy sets on two-dimensional data was high, reducing the classification difficulties caused by mixed samples. After comparing and analyzing different algorithms, the proposed transfer learning linear-discriminant analysis-axiomatic fuzzy set algorithm achieved the highest accuracy of 98.67% and completed data classification in about 20s, which was superior to other commonly used classification algorithms. The practical significance of the research lies in providing an efficient and accurate book sorting algorithm, which helps to improve the work efficiency and service quality of libraries.**

Keywords*—Linear discrimination; library intelligent services; book sorting; university libraries

I. Introduction

In the digital age, rapidly advancing information technology has not only completely changed learning and working methods, but also reshaped the service model of university libraries [1]. In today's era, university libraries are no longer just storage places for physical books. It has become a digital processing factory for knowledge and an interactive platform for cultural exchange [2]. With the integrated application of big data and intelligent systems, libraries can capture readers' reading habits and preferences in real-time, thereby providing more accurate and personalized services [3]. This transformation has made book management more intelligent, simplifying the process of book classification, retrieval, and borrowing, and greatly improving the user experience [4]. As a key link in intelligent library services, the improvement of automation and intelligence levels in book sorting is particularly important [5]. The current classification of books in libraries mostly relies on manual classification, which not only has low efficiency but also cannot classify books based on their content [6]. Instead, manual classification relies more on book names, which can lead to discrepancies between book categories and actual content. At present, the intelligent book sorting model still has the problem of low classification accuracy in practical applications. The inconsistency between book content and its partition can lead to users' distrust of the library, which is not conducive to the accumulation of users in the library management. Therefore, exploring and applying new intelligent book-sorting methods has become a research hotspot for library service innovation. In order to improve the efficiency and accuracy of book classification in libraries, it is proposed to use linear discriminant analysis technology to mine and classify the feature information of book content. At the same time, a dynamic learning framework was designed to learn the characteristic information of new book types in the library, in order to expand the book classification level of the system.

The study innovatively applies LDA technology to deeply explore the characteristic information of library collections. Based on the basic characteristics of library collections, a user-friendly book sorting assistant tool is designed, which helps to improve the service efficiency of library staff. The main contribution is to develop a standardized book classification evaluation system, which is convenient for other university libraries to reference and learn from, improving the intelligent service level of the entire industry.

The main content of the research is divided into six sections. Section I is an introduction to the significance and purpose of the research. Section II is a survey of related work, which elaborates on the limitations and shortcomings of the current research direction. Section III is a study on the construction of a library book sorting model designed for research. This section is divided into two parts. The first part is the processing method of library book feature information, and the second part is the construction of a book sorting model based on kilometer fuzzy sets. Section IV of the manuscript presents the main results of the research, including training and testing of the constructed model, as well as practical application validation. Section V is a discussion of the research results, comparing the effectiveness of the proposed model with other models in the current field. Section VI is a summary of the main research content.

II. Related Works

LDA is a common statistical analysis method. Ricciardi C et al. proposed a data mining method to analyze the characteristics of patients with myocardial ischemia. 22 features were extracted. The results showed that this method

could assist clinical decision-making. The principal component analysis could reduce the number of features [7]. Ontstick L L et al. conducted a double-blind, randomized, controlled clinical trial to identify risk factors associated with mortality in Ebola patients and evaluate the predictive ability of both methods. The results showed that binary logistic regression and LDA selected the same risk factors, predicted similar mortality rates, and scored similar areas under the curve [8]. Bonati L et al. proposed descriptors to characterize these states to obtain appropriate collective variables from metastable information. Neural networks compressed information in low dimensional space and used Fisher LDA as the objective function to maximize the network's discriminative power. This method was tested on alanine dipeptides and nonlinear separable datasets. An inter-molecular aldol condensation reaction was studied [9]. Yan et al. proposed an improved LDA to address the edge classes in multi class classification. The experimental results showed that the self-weighted LDA performed better than other methods, while also having higher computational efficiency [10]. Zhou L et al. conducted a cross-sectional study to investigate the gut microbiota and metabolic function in patients with polycystic ovary syndrome, recruiting participants from different groups. The LDA could effectively count the differences and classify them [11]. Ngailo E K et al. proposed a mis-classification probability approximation algorithm based on the LDA method to solve the mis-classification approximation problem in classification problems. The approximate mis-classification probability of the unknown covariance matrix was derived. The results showed that this method improved the approximation value of mis-classification probability [12].

Intelligent services in university libraries are a necessary condition for the development of university libraries in the information age. Mohammed SH et al. used topic modeling method to solve the classification problem of Chinese text data in digital libraries. Comparing the two popular methods of Latent Dirichlet Allocation and latent semantic allocation, the former performed better. When the topics were 20, the highest coherence of the potential Dirichlet allocation was 0.592179, while the potential semantic allocation was 0.577302 [13]. Asemi A et al. used descriptive and content review methods to review articles related to robotics in libraries and information science from 2007 to 2017. The intelligent systems contributed to multiple purposes for librarians, including technical services, public services, and consultation desks. Through integration with artificial intelligence technology, current information systems had high potential for improvement [14]. Wan M et al. proposed a cloud-based product service system platform to address the challenge of urban waste issues. This platform aimed to manage solid waste management resources. Case analysis verified the feasibility of the platform [15]. Chiu CK et al. proposed a learning status management system in an intelligent classroom to improve the learning status of students. Sensor technology and image recognition technology were used to detect and collect various information from students, and Bayesian classification networks were used to infer the learning status. In addition, the system also included a feedback mechanism. Two experiments verified the accuracy and effectiveness. The learning status analysis was highly consistent with observed results. Students were more focused in class [16].

The current research on book sorting still adopts the method of classifying books based on their semantic names, which leads to significant deviations in the accuracy of book collection and sorting. Moreover, the current intelligent management platform for library management still faces problems of inability to retrieve or inaccurate retrieval results in practical applications. Aiming at the shortcomings of the current intelligent management and sorting platform for library management, this study proposes the use of LDA technology to extract feature information of key content of books and sort them based on the key content features of books. The research has changed the main sorting basis of current book sorting technology, adopting a new sorting mode to achieve the goal of improving the accuracy of book sorting.

In summary, LDA is a commonly used statistical method that has been frequently applied in various studies. Influenced by information technology, university libraries are no longer limited to collecting images and are beginning to focus on intelligent service systems. Classification management technology has rich applications in various fields. To improve the service quality, the intelligent service system of university libraries is utilized to achieve the classification management of books. Therefore, the study combines LDA with intelligent services in university libraries to conduct book-sorting research.

III. Book Sorting Model based on LDA and AFS

With the development of the information age, university libraries have also begun to develop intelligent services. Intelligent book sorting can effectively reduce library management costs and improve library classification accuracy. In Section II, LDA and AFS are used to study the book sorting methods in university libraries. A book sorting model is constructed. The research will be conducted from two aspects. The first is data dimensionality reduction based on linear discrimination, and the second is the image book sorting model based on AFS.

A. Data Dimensionality Reduction Based on Linear Discriminant

LDA is a classic statistical method widely applied in pattern recognition and machine learning. It is a supervised learning algorithm mainly used for dimensionality reduction and classification problems [17]. The core idea is to project high-dimensional data onto a low-dimensional space that best distinguishes categories. Specifically, LDA attempts to find a series of projection axes such that when data is projected onto these axes. The distance between different categories is as large as possible, and the data dispersion within the same category is as small as possible, as shown in Fig. 1 [18].

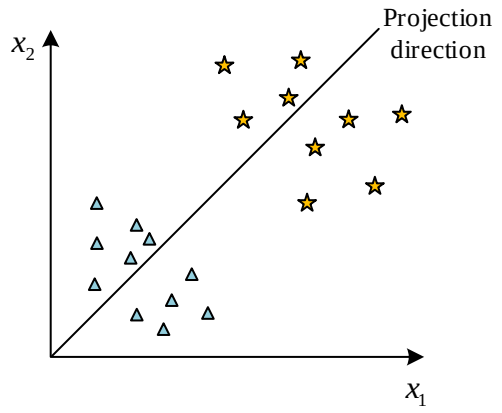


Fig. 1. LDA.

In the LDA algorithm, it is assumed that the data sample set is D. This sample set contains multiple data points. Each data point x_i can be considered as a vector in an n-dimensional space. Each component of this vector represents a feature of the data. Each data point x_i usually has an associated label category y_i, which indicates the classification to which the data point belongs. These categories are crucial for establishing classification models and making predictions in the future. In each category, it is assumed to be a classification labeled c. These classifications all have a corresponding sample size N_c. The number of samples determines the representativeness of each classification in the sample set. Generally speaking, a larger sample size indicates that the statistical characteristics of the corresponding classification are more reliable. To further analyze each category, the set of sample points for each data category is defined as X_c. For each category c, two important statistical measures can be further defined: the mean vector μ_c and the covariance matrix Σ_c. The mean vector is shown in Eq. (1) [19].

$$\mu_c = \frac{1}{N}\sum_{x\in X_c} x, c = (0,1) \tag{1}$$

The covariance matrix is shown in Eq. (2).

$$\Sigma = \sum_{x\in X_c} (x-\mu_c)(x-\mu_c)^T, c = (0,1) \tag{2}$$

A common method when dealing with binary classification problems is to find an appropriate mapping that reduces the dimensionality of high-dimensional data sample points to one-dimensional space to simplify the classification process. In mathematics, this can be achieved by defining a normal vector w, which essentially describes a vector in a one-dimensional spatial direction. Through this normal vector, the projection of each data sample point x_i on the line is represented as $w^T x_i$. Similarly, the mean values of different categories of sample points in the dataset can also be projected onto this vector to generate new mean points $w^T\mu_0$ and $w^T\mu_1$. Based on these criteria, LDA finds the optimal projection direction w through a mathematical optimization process. The original binary classification problem is transformed into a simple one-dimensional decision problem. In this way, no matter how high the dimensions of the original data are, theoretically, it can be effectively classified. The optimization objective is shown in Eq. (3).

$$MaxJ(w) = \frac{\left\|w^T\mu_0 - w^T\mu_1\right\|_2^2}{w^T\sum_0 w + w^T\sum_1 w} = \frac{w^T(\mu_0-\mu_1)(\mu_0-\mu_1)^T w}{w^T\left(\sum_0 w + w\sum_1 w\right)w} \tag{3}$$

In Eq. (3), $MaxJ(w)$ refers to the distance from the sample point to the center point. If $(\mu_0-\mu_1)(\mu_0-\mu_1)^T$ is defined as s_b and $\sum_0 w + w\sum_1 w$ is defined as s_w, then it can be expressed as Eq. (4).

$$s_w = \sum_{x\in X_c}(x-\mu_0)(x-\mu_0)^T + \sum_{x\in X_c}(x-\mu_1)(x-\mu_1)^T \tag{4}$$

By combining Eq. (4) with Eq. (3) and simplifying them, Eq. (5) can be obtained.

$$MaxJ(w) = \frac{w^T s_b w}{w^T s_w w} \tag{5}$$

Eq. (5) is the generalized Rayleigh quotient. In practical applications, most data samples are two or more categories. In this case, the optimization objective of LDA sample classification can be rewritten as Eq. (6).

$$\frac{W^T S_b W}{W^T S_w W} \tag{6}$$

In Eq. (6), S_b refers to the inter class discrete matrix of the sample points. S_w refers to the intra class discrete matrix of the data sample. S_b is shown in Eq. (7).

$$S_b = \sum_{c=1}^{k} N_c(\mu_c-\mu)(\mu_c-\mu)^T \tag{7}$$

S_w is displayed in Eq. (8).

$$S_w = \sum_{c=1}^{k}\sum_{x\in X_c}(x-\mu_c)(x-\mu_c)^T \tag{8}$$

From Eq. (7) and Eq. (8), the optimization objectives of two-class LDA cannot optimize multi-class LDA objectives. Therefore, other optimization methods are adopted to optimize multi-class LDA. The optimization objective function of multi-class LDA can be defined, as displayed in Eq. (9).

$$MaxJ(W) = \frac{\prod_{diag} W^T S_b W}{\prod_{diag} W^T S_b W} \tag{9}$$

In Eq. (9), $\prod_{diag} W^T S_b W$ refers to the diagonal element product. W represents a matrix. The optimization model for

multi-class LDA can be transformed into Eq. (10).

$$MaxJ(W) = \prod_{i-1}^{d} \frac{W_i^T S_b W_i}{W_i^T S_b W_i} \quad (10)$$

LDA is a common data dimensionality reduction method. When LDA performs data dimensionality reduction, it needs to calculate the intra class and inter class divergence matrix, and matrix $S_W^{-1}S_b$. When calculating $S_W^{-1}S_b$, a Lagrangian function needs to be introduced. The maximum eigenvalue $S_W^{-1}S_b$ and corresponding eigenvector w can be calculated. Fig. 2 displays the specific process [20].

Manifold Learning is a type of machine learning method used for dimensionality reduction, aimed at discovering and utilizing the underlying structure of the data itself to reveal its intrinsic features. These methods are based on the manifold assumption, which means that high-dimensional data is actually distributed on an unknown low-dimensional manifold. This low-dimensional manifold is implant in a high-dimensional space. The implementation of manifold learning is shown in Fig. 3 [21].

The manifold learning is applied to mine the potential structures of data in high-dimensional data spaces and obtain low-dimensional manifold representations. Then, through LDA, the low-dimensional manifold is mapped to a more discriminative space. The process of constructed Transfer Learning-LDA (TL-LDA) algorithm is shown in Fig. 4.

B. Book Sorting based on AFS

A fuzzy set is a mathematical theory used to handle uncertainty and fuzzy information, which can effectively deal with fuzzy and imprecise problems. The Membership Function (MF) of fuzzy sets can be used to describe the characteristics and attributes of fuzzy sets. AFS is a form of fuzzy set theory that satisfies a specific set of axioms and operational rules. These axioms and rules define the basic properties and operation methods of fuzzy sets, making the operations and reasoning of fuzzy sets consistent and reliable. AFS can model, represent, and process fuzzy concepts to solve real-world fuzzy problems. The research content of AFS includes membership functions of fuzzy sets, operation rules, feature descriptions, and inference methods, as shown in Fig. 5.

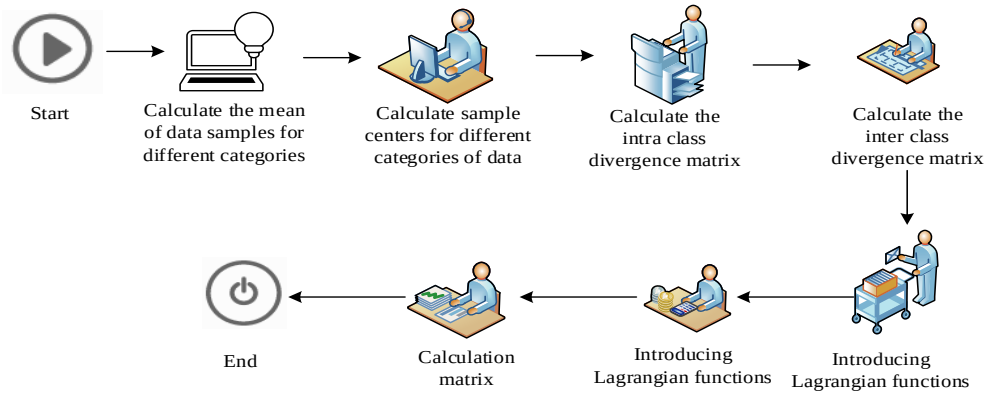


Fig. 2. LDA flow path.

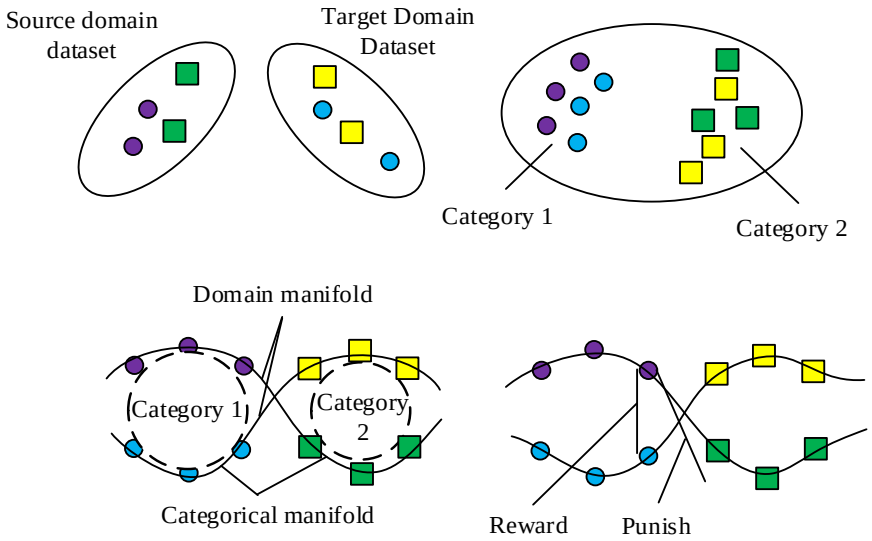


Fig. 3. The manifold learning framework.

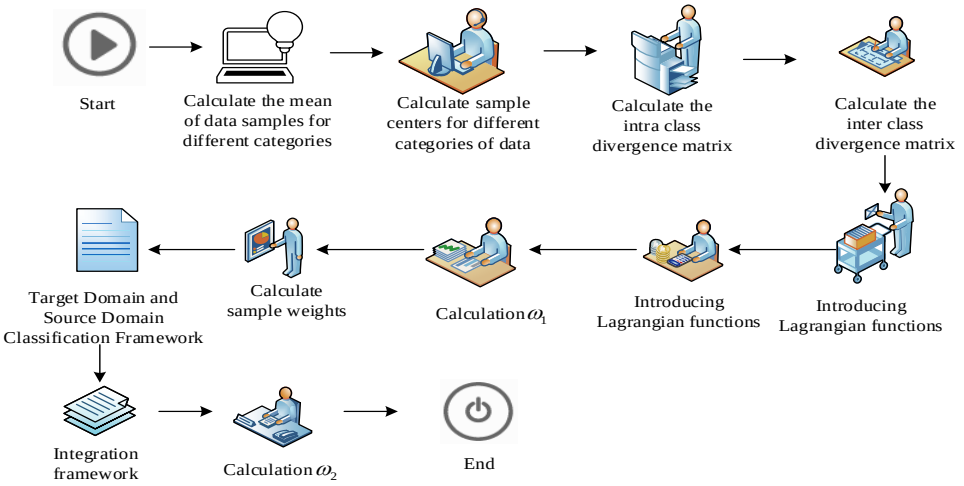


Fig. 4. Transfer learning - LDA algorithm process.

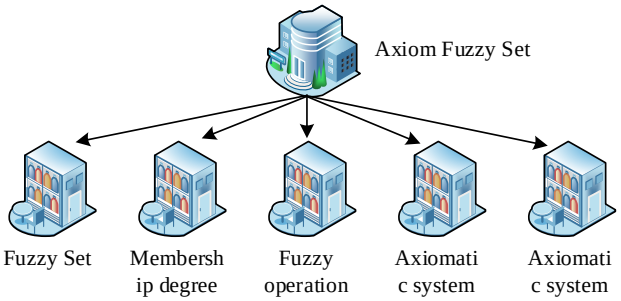


Fig. 5. Axiomatic fuzzy set of basic concepts.

Fuzzy sets refer to traditional set theory where an element either belongs to a set or not. In fuzzy set theory, the MF concept is introduced to represent the degree to which an element belongs to a certain set, which is represented by a numerical value in [0, 1]. Membership degree is the degree to which an element corresponds to a fuzzy set, reflecting the agreement between that element and the features of the fuzzy set. The MF is applied to determine the membership degree of an element in a fuzzy set. Fuzzy operations include fuzzy intersection, union and complement, which correspond to the traditional set theory. However, in their definition and operation process, the membership degree of the elements is considered. Axiom systems are designed to construct a robust fuzzy logic system. The AFS method is based on a set of strictly defined axioms that describe the basic rules that membership functions should follow, ensuring theoretical consistency and closure. Approximation reasoning refers to the reasoning process in fuzzy systems based on known information and axiomatic rules for unknown or uncertain situations. If M is assumed to be a simple set of concepts on Ω, (Ω, F, P) represents the probability measure space, and the observation dataset of (Ω, F, P) is represented as $X \subset \Omega$, (M, τ, Ω), and $(M, \tau | x, X)$. ξ is a concept that conforms to Eq. (11).

$$\xi = \sum_{i \in I} \left(\prod_{m \in A} m \right) \in EM \tag{11}$$

In Eq. (11), M represents a simple concept, which belongs to the set A. The membership function of AFS can be defined as Eq. (12).

$$\mu\xi(x) = \sup_{i \in I} \inf_{\gamma \in A_i} \frac{\sum_{u \in A_i^r} \rho_\gamma(u) N_u}{\sum_{u \in X} \rho_\gamma(u) N_u}, \forall x \in X \tag{12}$$

In Eq. (12), N_u represents the number of times sample $x \in X$ has been observed in the dataset. ρ_γ represents the weight function of a simple concept. $\rho_\gamma(x)$ is continuous on $\gamma(x)$. X is a set of randomly obtained data in probability space, and $\gamma \in M$. Then for $\forall x \in \Omega$, when $|X|$ approaches infinity, Eq. (12) can converge to Eq. (13).

$$\mu\xi(x) = \sup_{i \in I} \inf_{\gamma \in A_i} \frac{\int_{A_i^r(x)} \rho_\gamma(t) dP(t)}{\int_\Omega \rho_\gamma(t) dP(t)}, \forall x \in \Omega \tag{13}$$

After determining the membership function of the AFS structure, a book sorting model is constructed. The model uses TL-LDA to reduce the dimensionality of book data, extract the corresponding features of these data, and calculate the weight ratio of each data. AFS calculates the eligibility of each data, obtains a brief explanation, and then completes the corresponding data classification based on the brief explanation. The specific process is shown in Fig. 6.

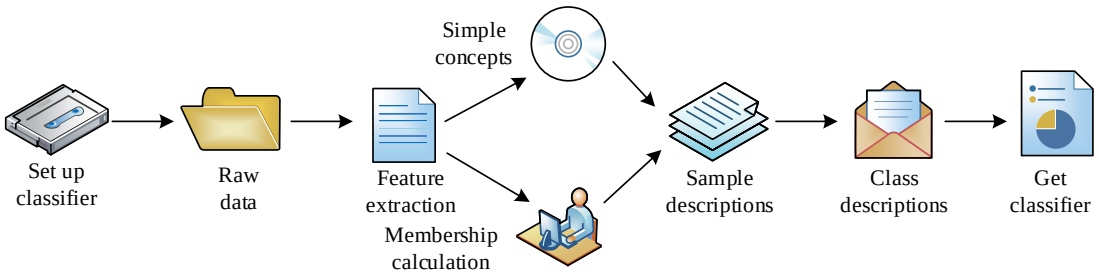


Fig. 6. Process of book sorting algorithm based on TL-LDA-AFS.

IV. Simulation Experimental Analysis of Book Sorting Model Based on LDA and AFS

In Section II, a book sorting model based on LDA and AFS is constructed. To verify the effectiveness, simulation experiments and analysis are designed on the model in Section III. Section III is divided into two parts. The first part sets the experimental environment and parameters, and the second part analyzes the simulation experiment results.

A. Experimental Environment and Parameter Settings

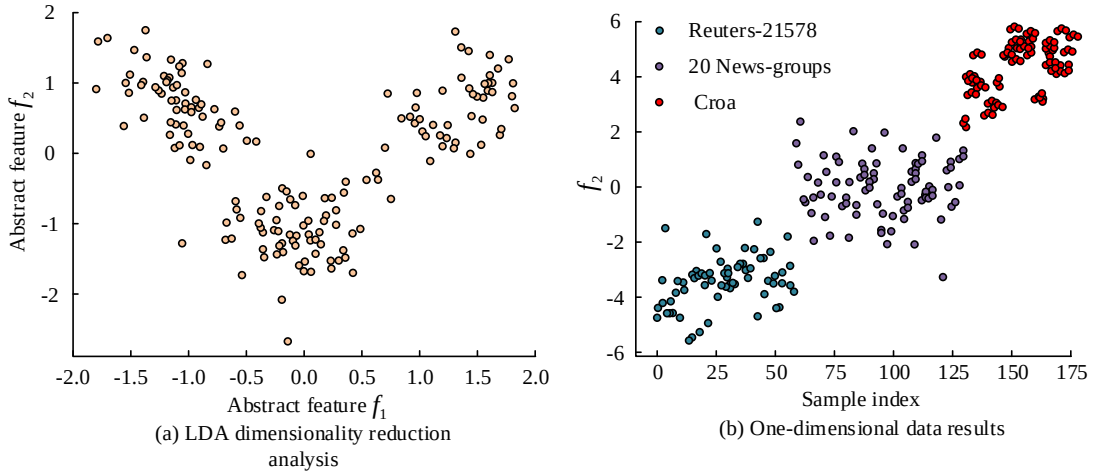
The experimental environment for the study is built in Python. The operating system is Windows 7 64bit, the processor is Inter (R) Core i5-12440, and the memory size is 16GB. The dataset used in the study includes the Reuters-21578 dataset, the 20 News-groups, and the Cora. Reuters-21578 is an iconic dataset in text classification, consisting of 21578 documents that provide detailed records of various major events and industry dynamics that have occurred globally. 20 The 20 News-groups is a large-scale text dataset designed specifically for research in text mining and text classification. The Cora dataset is an online document library containing approximately 37000 detailed computer science research papers in computer science research. These three datasets are used to generate 12 pairs of datasets as training and detection datasets for the model. The data ratio between the training set and the test set is 7:1, as displayed in Table I.

B. Data Dimensionality Reduction Processing Analysis

The TL-LDA performs LDA to reduce the dimensionality of the training set data. The results are shown in Fig. 7.

TABLE I. Experimental Dataset

Data number	Training task		Detection task	
	Positive	Negative	Positive	Negative
1	Orgs (1)	People (1)	Orgs (2)	People (2)
2	Place (1)	poisonous(enlarging)	Place (2)	Poisonous (tapering)
3	People (1)	edible(enlarging)	People (2)	edible(tapering)
4	encryption	protocols	compression	routing
5	protocols	Probabilistic methods	routing	genetic algorithms
6	encryption	probabilistic methods	compression	algorithms
7	comp.os	sci.crypt	comp.mac	sci.space
8	sci.med	rec.sport.baseball	talk.politics.gun	rec.sport.hokey
9	talk.religion	talk.religion	sci.crypt	talk.politics.guns
10	Computation-complexity	encryption	Computation-geometry	compression
11	Computation-complexity	probabilistic methods	Computation-geometry	genetic algorithms
12	Computation-complexity	protocols	Computation-geometry	routing



(a) LDA dimensionality reduction analysis

(b) One-dimensional data results

Fig. 7. LDA dimension reduction.

Fig. 7(a) shows the results of LDA dimensionality reduction on the training set data. The main data was concentrated in three regions. Fig. 7(b) shows the results of three datasets after reducing the training dataset to 1 dimension. After the data is reduced to 1 dimension, the three basic data exhibited a linear distribution. Samples of different categories may be confused within certain value ranges. The membership degree of these confused samples in specific environments is analyzed in detail to determine the MF. After the data is reduced to 1 dimension, the number of data features decreases, making it difficult for the model to distinguish these confusing samples. Therefore, reducing the dimensionality to 2D data is sufficient when performing dimensionality reduction processing. In situations with more features, more features can be applied to create fuzzy sets, which enhances classification ability. When performing discrimination, AFS can make judgments based on the membership degree on different concepts, further improving the universality and accuracy. However, it's not that having more features leads to better performance. Excessive features may lead to the curse of dimensionality in fuzzy set rules. In addition to the difficulty of model parsing, this may also consume more computing resources. Therefore, balancing the number of features is crucial for building an effective fuzzy set model. To further determine the degree of pre-processing, the accuracy between one-dimensional data and two-dimensional data is compared. The results are shown in Fig. 8.

In Fig. 8, regardless of whether the data was reduced to one-dimensional or two-dimensional, the accuracy in extracting data features continuously decreased with the increase of the ratio. When the ratio was 0.8, the accuracy of feature extraction for both dimensions decreased to the lowest. The accuracy of feature extraction for one-dimensional data was only 0.3248, while the two-dimensional data was 0.4676, which was 0.1428 higher than the one-dimensional data. When the ratio was 0.1, the accuracy of feature extraction in both dimensions increased to the highest. The accuracy of feature extraction for one-dimensional data was 0.5248, and the two-dimensional data was 0.6402, which was 0.1154 higher than that of one-dimensional data. The feature extraction accuracy of two-dimensional data was equal before the ratio was less than 0.4. The accuracy of feature extraction in one-dimensional data was always changing. When the ratio was 0.5, the accuracy of feature extraction in two-dimensional data was 0.5964, while the one-dimensional data was only 0.4856, which was 0.1008 lower than that in two-dimensional data.

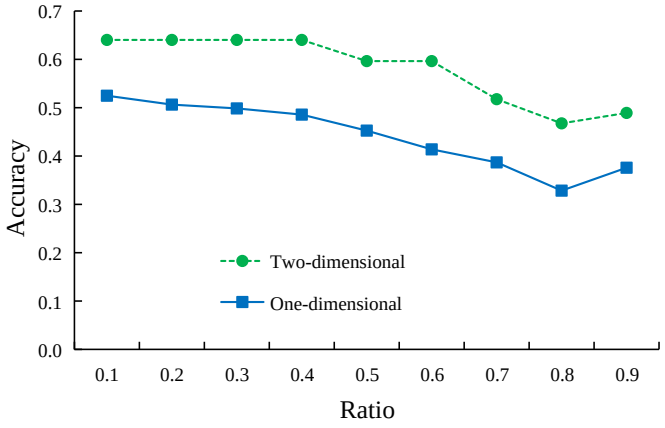


Fig. 8. Comparison of the feature extraction accuracy rate.

C. Analysis of Classification Results

After the dataset is dimensionally reduced by LDA, the membership degree is calculated using the reduced data. The results are shown in Fig. 9.

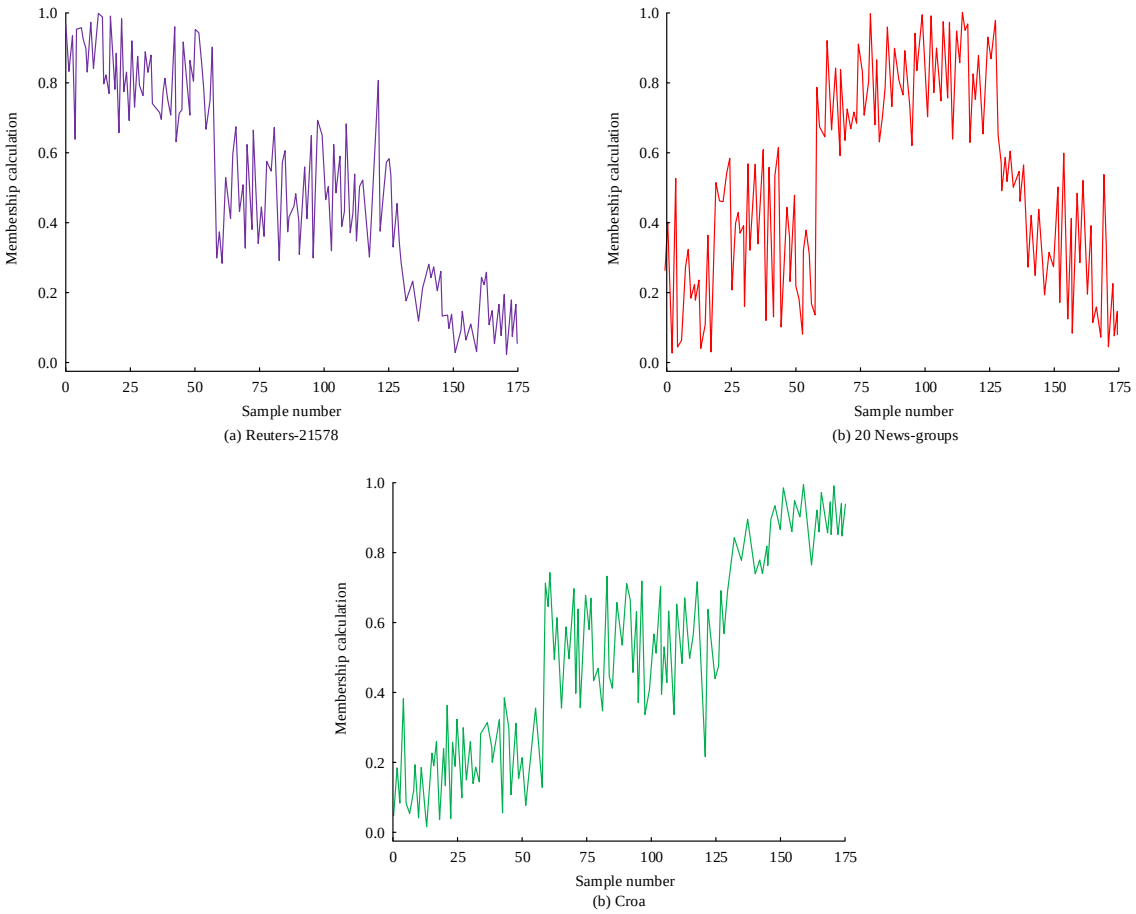


Fig. 9. Membership levels on the different datasets.

Fig. 9(a) shows the membership degree on the Reuters-21578 dataset. The membership degree of the top 50 samples was concentrated between 0.6 and 1.0. The membership degree of samples 50 to 125 was concentrated between 0.2 and 0.6. Samples 125 to 175 were concentrated between 0.0 and 0.3. Fig. 9(b) shows the membership degree on the 20 News-groups dataset. The membership degree of the top 50 samples was concentrated between 0.0 and 0.6. Samples 50 to 125 were concentrated between 0.6 and 1.0. Samples 125 to 175 were concentrated between 0.0 and 0.6. Fig. 9(c) shows the membership degree on the Cora dataset. The membership degree of the top 50 samples was concentrated between 0.0 and 0.4. Samples 50 to 125 were concentrated between 0.2 and 0.8. Samples 125 to 175 were concentrated between 0.8-1.0. The study describes the membership degrees of three datasets, as shown in Fig. 10.

Fig. 10 shows the membership degree distribution of three categories relative to all samples in the same chart. It is possible to visually observe the membership values of each category on different samples. The significant differences in membership indicate that classification decisions can be made based on these differences, thereby improving the accuracy and efficiency of classification. There are significant differences in the membership degrees of the three different datasets across all samples. Support Vector Machine (SVM), K-nearest Neighbors (KNN), Native Bayes (NB), and Logistic algorithms are commonly used classification algorithms. To verify the feasibility of TL-LDA-AFS, the accuracy and classification time of these algorithms are compared on the dataset. The results are shown in Fig. 11.

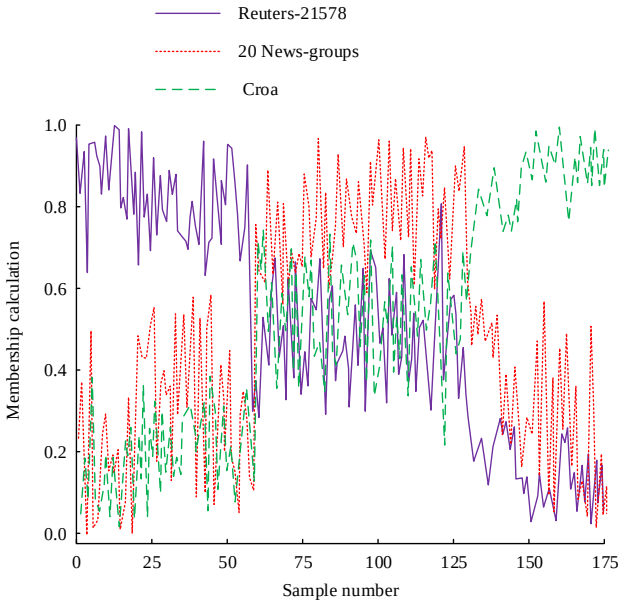
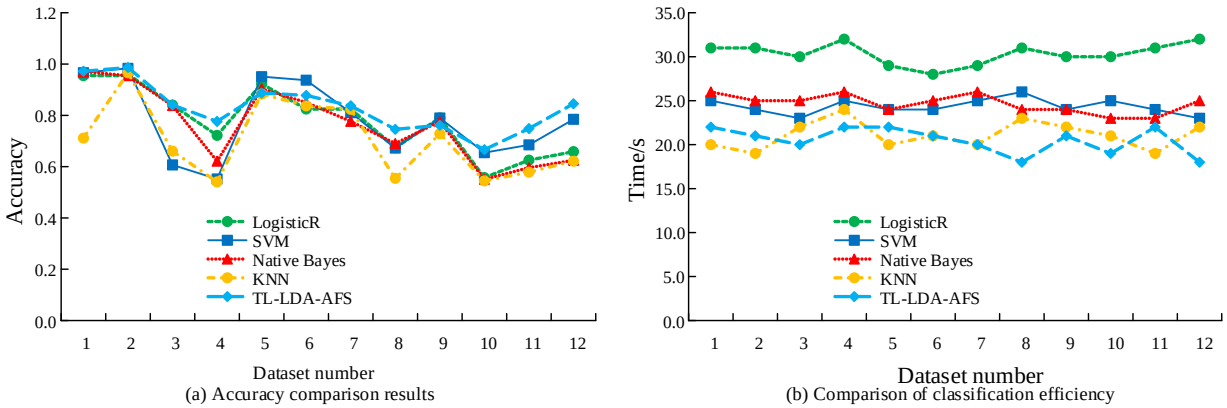


Fig. 10. Class description membership comparison.



(a) Accuracy comparison results (b) Comparison of classification efficiency

Fig. 11. Model feasibility validation.

Fig. 11(a) shows the accuracy comparison. In dataset 2, all algorithms achieved the highest accuracy. The accuracy of SVM, KNN, NB, Logistic, and TL-LDA-AFS algorithms were 0.9833, 0.9660, 0.9547, 0.9600, and 0.9867, respectively. In dataset 10, all algorithms had the lowest accuracy. The accuracy of SVM, KNN, NB, Logistic, and TL-LDA-AFS algorithms were 0.6545, 0.5445, 0.5507, 0.5572, and 0.6662, respectively. Fig. 11(b) shows the algorithm time consumption. On all datasets, the TL-LDA-AFS algorithm had the lowest classification time, taking about 20s to complete the data classification. SVM, KNN, NB, and Logistic took 24s, 20 seconds, 25s, and 30s to complete data classification, respectively. Whether it is the accuracy of data classification or classification efficiency, the TL-LDA-AFS algorithm has higher accuracy. In text datasets, its performance is significantly better than other classification algorithms.

V. Discussion

A study has proposed an intelligent book classification method based on LDA and AFS to improve the efficiency and accuracy of book classification in university libraries. Asemi et al. [14] found through content review methods that intelligent systems contribute to multiple aspects of libraries. Therefore, the study adopts artificial intelligence technology to design a book sorting model. Mohammed et al. [13] used topic modeling methods to solve the classification problem of Chinese text data in digital libraries and found that latent Dirichlet allocation performed better than latent semantic allocation. Similar to the research findings, the use of artificial intelligence technology can effectively improve the level of book sorting.

The experimental results show that when the data is reduced to two-dimensional, the accuracy of feature extraction reaches 64.02%, which is significantly higher than the 52.48% of one-dimensional data. The transfer learning linear discriminant analysis axiomatic fuzzy set (TL-LDA-AFS) algorithm proposed in the study achieved the highest classification accuracy of 98.67% among all comparison algorithms, and completed data classification in approximately 20 seconds, significantly better than other commonly used classification algorithms. The intelligent book classification method based on LDA and AFS proposed in the study not only improves the accuracy and efficiency of book classification, but also provides a new intelligent solution for book management in university libraries.

VI. Conclusion

The research aims to create an intelligent book sorting system for university libraries to improve book management efficiency and user experience. Based on the LDA method, this study successfully reduces the dimensionality of multidimensional feature data to two-dimensional, making it easier to classify books more accurately. From the experimental results, when the data was reduced to 1 dimension, the three basic data exhibited a linear distribution. When the ratio was 0.8, the accuracy of feature extraction for both dimensions decreased to the lowest. The accuracy of feature extraction for one-dimensional data was only 0.3248, while the accuracy for two-dimensional data was 0.4676, which was 0.1428 higher than the one-dimensional data. The accuracy of feature extraction for LDA processed data in two-dimensional dimensionality reduction was 64.02%, which was approximately 11.54% higher than that of one-dimensional data. After comparing different classification algorithms, the TL-LDA-AFS algorithm developed in this study achieved the highest accuracy of 98.67% on dataset 2, which was the highest among all comparison algorithms. It also showed significant advantages in classification efficiency, taking only about 20s to complete classification, greatly improving the intelligence level of book sorting in university libraries. The study provides a fast and accurate intelligent book sorting algorithm for university libraries. It is expected to promote the progress of library management towards intelligence and automation, while improving the quality and efficiency of library services for teachers and students. The book sorting model designed for research has a large amount of data computation and high system load requirements when learning the feature information of different categories of books. It takes a long time to process the newly added book categories in the library, resulting in a high resource load on the book sorting system. Future research will further optimize the model's ability to learn feature information for new types of books, reduce the model's learning time for new classifications, and lower the model's burden on system resources.

REFERENCES

[1] F. Zhang, "Research on Keyword Mining of Academic Library Subject Service," Acta Informatica Malaysia (AIM), vol. 7(2), pp. 97-100, 2023.

[2] C. Daraojimba, O. Onunka, T. Onunka, A. A. Fawole, I. J. Adeleke, "Library and Information Services in The Digital Age: Opportunities and Challenges," Acta Informatica Malaysia (AIM), vol. 7(2), pp. 113-121, 2023.

[3] Lund B D, Wang T. Chatting about ChatGPT: how may AI and GPT impact academia and libraries. Library Hi Tech News, 2023, 40(3): 26-29.

[4] Winata A P, Fadelina R, Basuki S. New normal and library services in Indonesia: A case study of university libraries. Digital Library Perspectives, 2021, 37(1): 77-84.

[5] Lee P C. Technological innovation in libraries. Library Hi Tech, 2021, 39(2): 574-601.

[6] Tella A. Robots are coming to the libraries: are librarians ready to accommodate them. Library Hi Tech News, 2020, 37(8): 13-17.

[7] Ricciardi C, Valente A S, Edmund K, Cantoni V, Green R, Fiorillo A, Cesarelli M. Linear discriminant analysis and principal component analysis to predict coronary artery disease. Health Informatics Journal, 2020, 26(3): 2181-2192.

[8] Ontshick L L, Sabue J C M, Kiangebeni P M, Mbayi O T, Ntamabyaliro J M N, Tamfumu J J M, Matendo R M M. Comparison of the performance of linear discriminant analysis and binary logistic regression applied to risk factors for mortality in Ebola virus disease patients. Journal of Electronics, Electromedical Engineering, and Medical Informatics, 2023, 5(3): 205-210.

[9] Hasanvand M, Nooshyar M, Moharamkhani E, Selyari A. Machine Learning Methodology for Identifying Vehicles Using Image Processing. AIA, 2023, 1(3): 170-178.

[10] Yan C, Chang X, Luo M, Zheng Q, Zhang X, Li Z, Nie F. Self-weighted robust LDA for multiclass classification with edge classes. ACM Transactions on Intelligent Systems and Technology (TIST), 2020, 12(1): 1-19.

[11] Zhou L, Ni Z, Cheng W, Yu J, Sun S, Zhai D, Cai Z. Characteristic gut microbiota and predicted metabolic functions in women with PCOS. Endocrine Connections, 2020, 9(1): 63-73.

[12] Ngailo E K, Chuma F. Approximation of misclassification probabilities in linear discriminant analysis based on repeated measurements. Communications in Statistics-Theory and Methods, 2023, 52(23): 8388-8407.

[13] Mohammed S H, Al-augby S. Lsa & lda topic modeling classification: Comparison study on e-books. Indonesian Journal of Electrical Engineering and Computer Science, 2020, 19(1): 353-362.

[14] Asemi A, Ko A, Nowkarizi M. Intelligent libraries: a review on expert systems, artificial intelligence, and robot. Library Hi Tech, 2020, 39(2): 412-434.

[15] Wan M, Qu T, Huang M, Li L, Huang G Q. Cloud-based product-service systems platform for household solid waste classification management. IET Collaborative Intelligent Manufacturing, 2020, 2(2): 66-73.

[16] Chiu C K, Tseng J C R. A bayesian classification network-based learning status management system in an intelligent classroom. Educational Technology & Society, 2021, 24(3): 256-267.

[17] Kim S, Rigatto K, Gazzana M B, Knorst M M, Richards E M, Pepine C J, Raizada M K. Altered gut microbiome profile in patients with pulmonary arterial hypertension. Hypertension, 2020, 75(4): 1063-1071.

[18] Xu C, Shen J. Multi-criteria decision making and pattern recognition based on similarity measures for Fermatean fuzzy sets. Journal of Intelligent & Fuzzy Systems, 2021, 41(6): 5847-5863.

[19] Bhattacharyya R, Mukherjee S. Fuzzy membership function evaluation by non-linear regression: An algorithmic approach. Fuzzy Information and Engineering, 2020, 12(4): 412-434.

[20] Lazaroiu G C, Roscia M. Fuzzy logic strategy for priority control of electric vehicle charging. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 19236-19245.

[21] Jan N, Gwak J, Pei J, Maqsood R, Nasir A. Analysis of networks and digital systems by using the novel technique based on complex fuzzy soft information. IEEE Transactions on Consumer Electronics, 2022, 69(2): 183-193.

Prediction of Booking Trends and Customer Demand in the Tourism and Hospitality Sector Using AI-Based Models

A Case Study of Major Hotel Chain

Siham Rekiek[1]*, Hakim Jebari[2], Kamal Reklaoui[3]
Intelligent Automation and BioMedGenomics Laboratory, University Abdelmalek Essaâdi, Morocco[1]
Innovative Systems Engineering Laboratory, University Abdelmalek Essaâdi, Morocco[2]
Innovation and Engineering Research Laboratory, University Abdelmalek Essaâdi, Morocco[2, 3]

***Abstract*—Accurate demand forecasting is critical for optimizing operations in the tourism and hospitality sectors. This paper proposes a robust multi-algorithmic framework leveraging four advanced models of Artificial Intelligence (LSTM, Random Forest, XGBoost, and Prophet) to predict booking trends and customer demand. In contrast to traditional approaches, this study incorporates external factors such as competitors' pricing strategies, local events, and weather patterns, offering a more holistic view of demand drivers. Using a comprehensive dataset from a leading hotel chain, we systematically compare the performance of these models, providing detailed evaluations. The findings offer actionable insights for hotel managers, demonstrating how predictive analytics can inform revenue management, improve operational efficiency, and enhance marketing initiatives. These results contribute to the evolving field of demand forecasting, offering practical recommendations for data-driven decision-making in the tourism and the hospitality sector.**

***Keywords*—Artificial Intelligence; decision-making; long short-term memory; XGBoost; Random Forest; Prophet; tourism; hospitality; demand forecasting; booking trends; customer**

I. Introduction

The tourism and hospitality sector is one of the most important sectors in the world economy. The application of Artificial Intelligence has transformed this sector and other sectors [1, 2], and its support is used to enhance service quality and productivity. As more AI technologies such as machine learning, natural language processing and robotics become readily available [3], tourism and hospitality businesses can leverage them to innovate their business processes in marketing, customer service, inventory management or revenue optimization [4, 5].

Hotels, as well as other tourism service providers, face the predicament of obtaining an equilibrium between demand and supply. Overestimating the demand often lead to an overstaffing situation and waste of resources. Conversely, underestimating demand can cause a high opportunity cost of unattended customer demands and potential loss in revenue due to lack of supply. Traditional methods were designed for forecasts in those conditions where this information is close enough such as mid-term city-wide events; however, when confronted with short-term rare event related increases at a hotel caused by conferences, festivals or sudden extreme weather changes these approaches quickly outperformed given the need to incorporate variables with diverse scales as well as volatile derivatives along nonlinearity. Artificial Intelligence introduces flexible modeling that provides this improvement on classical estimation models allowing one to learn from more different sources because it can consider many exogenous factors simultaneously while capturing nonlinear relations among them.

The main problem addressed by this paper is the challenge of accurately predicting customer demand in the tourism and hospitality industry using traditional methods, which fail to account for the complexity and volatility of real-world scenarios. Hotels often experience difficulties in adjusting to sudden demand changes, leading to either overstaffing or missed revenue opportunities. The inadequacy of current forecasting approaches to include critical external factors such as competitors' pricing, local events, and unpredictable weather patterns further compounds the issue. Therefore, there is a need to develop and evaluate AI-based multi-algorithmic models capable of addressing these complexities.

This research aims to answer the following key questions:

- How can AI-based multi-algorithmic models improve the accuracy of demand forecasting in the tourism and hospitality sector?
- What is the comparative performance of advanced AI techniques such as LSTM, Random Forest, XGBoost, and Prophet in predicting booking trends?
- How do external factors, like competitors' pricing, local events, and weather conditions, influence customer demand in the hospitality industry?
- What actionable insights can hotel managers derive from AI-powered demand forecasting to optimize revenue management and operational efficiency?

The objectives of this paper are:

- To develop a comprehensive multi-algorithmic forecasting framework that combines advanced AI

models (LSTM, Random Forest, XGBoost, and Prophet) for predicting booking trends and customer demand.

- To incorporate external factors such as competitors' pricing, local events, and weather patterns into the predictive models to improve forecast accuracy.
- To systematically compare the performance of these AI models using a real-world dataset from a leading hotel chain in Morocco.
- To provide practical recommendations for hotel managers on how to use predictive analytics for better decision-making in revenue management, operational efficiency, and marketing.

The significance of this research lies in its potential to revolutionize demand forecasting in the tourism and hospitality sector through the use of AI. By developing more accurate and responsive forecasting models, this study can help hotel managers optimize their resource allocation, avoid overstaffing or under-booking, and enhance overall profitability. Furthermore, the integration of rarely considered external factors into the predictive models addresses the limitations of traditional approaches, making this research highly relevant for both academics and industry practitioners. The findings from this research can also be extended to other industries, such as retail and transportation, where demand forecasting plays a crucial role.

The paper is structured as follows: Section II provides a literature review, Section III presents the models and experimental setup, Section IV outlines the experimentation and results, Section V offers a discussion, and Section VI details future research directions, with the conclusion presented in Section VII.

II. Literature Review

A. Artificial Intelligence Technologies in Tourism and Hospitality

Artificial Intelligence technologies (machine learning, natural language processing, robotics, computer vision, augmented reality and blockchain) are increasingly employed within tourism and hospitality to produce service customer and efficient operations [5, 6]:

1) Machine learning is an algorithm able to learn from and provide recommendations or predictions depending on data is machine learning. Utilization of machine learning in tourism as well as hospitality is numerous. It can be used to predict customer preferences [7], to dynamically optimize the pricing of hotel rooms [8] as well as to optimize marketing efforts. Machine learning algorithms can analyze booking patterns and generate, based on these data, personalized travel recommendations for customers [9]. In addition, machine learning can also support inventory decision-making by demand forecasting and adjusting stocks when required [10].

2) Natural language processing is a subfield of Artificial intelligence which is focused on the interaction between computers and humans using the natural language. In tourism and hospitality, NLP is applied for chatbots and virtual assistants to answer instantly customers' questions or provide personalized recommendations [10, 11]. Furthermore, NLP helps analyzing customer feedbacks or reviews in order to extract information about customer satisfaction or points of weaknesses [12]. For instance, NLP algorithms are able to deal with huge amounts of unstructured data coming from social media or online reviews to extract trends and sentiment [13].

3) Robotics: Designing and operating machines for tasks which were previously done by humans is called robotics. Robots are used in the travel and hotel industry for chores like room cleaning, baggage handling, and check-in systems [13]. Robots may save labor expenses, improve operational efficiency, and provide visitors with an original and unforgettable experience [14, 15]. Robots with artificial intelligence, for instance, can negotiate difficult surroundings, communicate with visitors, and do highly exact jobs [16].

B. Emerging Artificial Intelligence Technologies

Beyond the well-established applications, several emerging Artificial Intelligence technologies hold promises for further transforming the tourism and hospitality sector such as Computer Vision, Voice Recognition, Augmented Reality and Blockchain:

1) Computer vision is an approach to artificial intelligence which permits machines to study and interpret visual input It is utilized in a variety of tourism-related applications. Hotels, for instance, are leveraging computer vision to ensure security by early detecting potential threats through surveillance systems [17] and facilitating automated check-ins and check-outs using facial recognition [18] as well as monitoring cleanliness and maintenance of hotel properties [19]. Computer vision is also applicable in the marketing context whereby it can capture customer demographics information as well as behaviours in retail spaces for effective product placement and promotions [20].

2) Voice recognition technology helps AI systems to understand and react to human voices. Recently, this technology is also been employed in hotel rooms. Guests can use their voice to control the room features such as lights, temperature or entertainment systems. Voice recognition promises higher convenience and guest personalization [19, 20]. Voice-activated assistants inform guests about services of the hotel, local restaurants etc., interesting spots and current weather [21]. Voice control may also help in speeding up check-in/check-out processes by reducing the time needed for filling paperwork at the front desk [22].

3) Augmented reality consists of enriching the real world with digital information. In the field of tourism, a visitor can be guided in an interactive way during his/her visit to a landmark or point of interest, or interact with virtual characters that will increase their knowledge about historic events or monuments. When booking a hotel room, customers may do a virtual tour in which they experience the look and

feel of the room through AR before taking a final decision [21, 22]. Similarly, when eating out at a restaurant, an AR system could provide a 3D visualisation of menu items and related dietary information on top of printed food menus or directly projected onto restaurant tables so as to help people make more informed choices [23]. AR has been used by destination managers to offer potential visitors immersive experiences advertising by letting them explore landscapes and other interest features without leaving home or office [24].

4) Blockchain technology which is known for secure and transparent nature can be used in multiple applications of tourism and hospitality. Blockchain can improve security and transparency of transactions, supply chain management, and veracity of customer reviews [25]. Blockchain technology may be used for safe creation of immutable transaction ledger that can reduce intermediaries, double selling frauds and ensure the integrity in booking and payment process [26]. Furthermore, blockchain could allows sharing data between various travel industry constituents securely [27].

C. Traditional Demand Forecasting Methods

Traditional demand forecasting techniques such as linear regression, ARIMA (Autoregressive Integrated Moving Average) and exponential smoothing have been used for a long time in tourism and hospitality sector. These techniques are easy to implement and interpret but lack the ability to capture the complexities among variables. For example, linear regression models assume that independent variables and demand have a linear relationship which may not be true in practical case-study environments. ARIMA models are better than linear regression but are not able to incorporate external factors like weather, events or competitor's prices. Many research papers have proved that traditional models fail to give promising results when it comes to accurately predicting hotel demand. The research in [28] states that ARIMA model provides inaccurate predictions when data exhibits high seasonality due to events as well as irregular patterns of demand especially during peak stays. The study in [29] suggests sophisticated forecasting methods with enhanced performances exhibiting non-linear patterns of multiple determinants.

D. Artificial Intelligence for Demand Forecasting

Artificial Intelligence (AI) has become a powerful tool to predict demand, especially in industries where demand is affected by multiple internal and external factors. Contrary to conventional models, AI techniques are capable of dealing with complex and non-linear relationships, while learning automatically from historical data. The most frequently used AI algorithms for demand prediction include the following.

1) Decision trees and Random Forests: These algorithms can be applied with any type of data (Numerical or Categorical), moreover it provides feature importance ranking that could help understand the impact of each factor on influencing the demand [30].

2) Support Vector Machines (SVM): SVM are efficient in high dimensional space and have been applied for hotel occupancy prediction as well. However, performance of SVM is parameter sensitive and scaling dependent [31].

3) Artificial Neural Networks (ANN): Artificial neural networks have been used for demand forecasting because of its ability to learn nonlinear and complex patterns, but the ANNs demand huge amounts of quality data and computational requirements [32].

4) Deep learning: LSTM is a recurrent neural network type [33] that is well-suited to time series forecasting problems due to their ability in handling long-term dependencies in data [34].

E. Influencing Factors in Tourism Demand Forecasting

Several studies emphasized the need to include external factors in forecast models, thus improving their prediction capacity.

1) Seasonality: The study in [35] investigates the seasonality of tourism and its implication on booking pattern. Hotels in popular tourist areas often face high demand during holidays and school breaks.

2) Competitor pricing: The study in [36] found that many customers compare prices on several platforms before booking, so competitor pricing is likely to have an impact on customer booking behaviour.

3) Local events: The study in [37] showed that large local events such as conferences, music festivals or sports tournaments can generate a peak of demand in the short-term so being able to integrate event-based information will greatly benefit forecasting.

4) Weather conditions: The research in [38] pointed out that the weather conditions significantly impact the hotel demand in leisure destination, especially for outdoor tourism activity.

III. Models and Experimental Setup

Based on the literature study and preliminary experiments, the following Artificial Intelligence models have been selected for further evaluation.

A. Long Short-Term Memory (LSTM)

LSTM, a special kind of recurrent neural network (RNN), is very popular in time series prediction since it can learn long-term and short-term dependency. LSTM networks maintain memory over time by using gated cells to control the flow of information.

1) Architecture: In this study, the LSTM model consists of 2 hidden layers each with 128 LSTM units, followed by a dense layer with single neuron for regression output. Dropout regularization was used to prevent overfitting.

2) Training: Adam optimizer with learning rate of 0.001 is used and mean squared error is taken as loss function. Early stopping based on validation performance is done after training for 50 epochs.

3) Feature selection: LSTM model uses lag features along with other booking related internal features, weather data and

few indicators to include external influence in the model and temporal patterns.

B. Random Forest

Random Forest is an ensemble learning method. Building multiple decision trees at training and outputting the average prediction of all the individual trees can reduce overfitting and improve model generalizability.

1) Feature importance: Random Forest provider a feature importance ranking which indicate are the most important variables driving booking demand. In this research room pricing, competitor rates, and local events were found to be the most important features.

2) Hyperparameters: The model was tuned by grid search with number of trees (n_estimators) set to 500, maximum depth set to 10 and minimum samples per leaf set to 5.

C. XGBoost

XGBoost is an optimized implementation of gradient boosting that will increase the speed and performance of the model. It builds decision trees one at a time, trying to reduce the errors from the last iteration. It also handles missing data.

1) Boosting rounds: The model was set to do 100 boosts/rounds with early stopping enabled so it doesn't overfit.

2) Regularization: L1 and L2 regularization terms are added to the objective function to prevent overfitting, especially when there are some noisy features such as competitor pricing data.

D. Prophet

Prophet is a time-series forecasting algorithm developed by Facebook to make fast and accurate predictions on non-linear trends, seasonality and holiday effects. It is very useful in the tourism industry because there are a lot of seasonal patterns and event-triggered peak or drop demand. Some of the key features are:

1) Seasonality components: You can easily model multiple seasonality components with Prophet (i.e. weekly and yearly) which makes it great for predicting hotel bookings as they have very clear seasonalities.

2) Holiday and event handling: In Prophet's model special event indicators are added to account for local festivals, conferences, etc, that cause demand to surge.

IV. Experimentation and Results

A. Data Collection and Preprocessing

The dataset used in this study is a real-world hotel booking dataset obtained from an internal data source of a well-known Major hotel chain in Morocco. The data collected ranged from 2015 to 2023. It includes:

1) Booking data: These captures booking date, room type, length of stay, booking channel and information on cancellations and no-shows.

2) Customer data: This captures nationalities, age, range and purposes of travel (business/leisure).

3) Pricing data: The room price at the time of booking, the discount applied to booking for prepayment or another reason, and variations in prices due to changes in demand, seasonality and other reasons.

4) Competitor data: Pricing data of competitor hotels fetched from 3rd party aggregators and web scraping.

5) External data: Weather conditions (Temperature, Precipitation), local event schedules (Conferences, Festivals).

External data such as weather information, and local events' data were also added to the dataset considering the impact of these factors on customer demand. Data preprocessing was performed in order to make sure that the dataset is suitable for analysis:

1) Handling missing data: Missing values in customer demographics were imputed with median or mode imputation. Missing weather or competitor pricing data were filled with interpolation.

2) Outlier detection and removal: Outliers in room rates, booking lengths and cancellations were detected using the Inter Quartile Range (IQR) method and capped within a reasonable range to avoid any distortion in the model.

3) Normalization and standardization: Continuous variables like room rates and customer ages were normalized using Min-Max scaling, while categorical variables like booking channels and room types were one-hot encoded.

B. Feature Engineering

Feature Engineering was an important factor to increase the predictive power of the Artificial Intelligence models. The following new features have been created:

1) Lag features: booking data from the last 7, 14 and 30 days before each reservation were used to create lag features so the model could account for trends and seasonality

2) Event indicators: Binary indicators have been constructed to identify the occurrence of major local events (festivals, sports tournaments, meetings etc.), which are expected to impact short-term demand positively.

3) Weather features: Features including temperature (in Celsius and Fahrenheit), humidity, and precipitation were updated, as the weather conditions have a significant differentiating impact in some specific leisure destinations in particular.

4) Price competitiveness: A variable which measures the difference between hotel's room rate and its competitor was defined to consider the effect of competitive pricing on demand.

C. Data Experimental Setup

The dataset was partitioned into training (80%) and testing (20%), datasets using data from 2015 to beginning 2021 as the training data and data from beginning 2022–early 2023 as testing dataset. To tune hyperparameters and validate the models, a 5-fold cross-validation step was used. After building

each model, their performance was further examined using both the training and testing datasets to evaluate their generalization capacity.

- Training Set: 2015-2021
- Testing Set: 2022-2023

D. Evaluation Metrics

Several performance metrics were used to assess the accuracy and reliability of the models:

- Mean Absolute Percentage Error (MAPE): A percentage-based measure of prediction accuracy, useful for comparing models across different scales.
- Root Mean Square Error (RMSE): A common measure of the magnitude of prediction errors.
- R-squared (R^2): Represents the proportion of variance in the dependent variable explained by the independent variables.

E. Results Analysis

The performance metrics of the models (LSTM, Random Forest, XGBoost, and Prophet) are summarized in Table I.

TABLE I. PERFORMANCE METRICS OF THE MODELS

Model	Metrics		
	MAPE (%)	***RMSE (%)***	***R^2***
LSTM	5.42	12.34	0.92
Random Forest	6.18	13.27	0.89
XGBoost	5.75	12.88	0.90
Prophet	7.12	14.22	0.87

The results show that:

- The LSTM model outperforms all models with the capability to capture both short-term and long-term demand dependencies. Based on the results, this model has an MAPE of 5.42% and $R^2 = 0.92$ which indicates highest accuracy among all developed models.
- Both Random Forests and XGBoost have relatively similar performance where Random Forests performs better in feature importance analysis while not delivering good performance in temporal prediction.
- Prophet was the weakest by far on both performance measures, not great at short-term fluctuations, but pretty comfortable with seasonality.

F. Feature Importance Analysis

Understanding the factors that affect customer demand is important before beginning any predictive modeling. This statement becomes more relevant and challenging in Tourism and Hospitality due to high level of competition as well as dynamic nature of sector. We used four different models Long Short-Term Memory, Random Forest, XGBoost and Prophet to understand which factors are affecting customer demand prediction the most. In such highly competitive industry, identifying these key features can help organizations allocate resources appropriately, make effective pricing decisions and create popular marketing campaigns. This section explores all such important features identified by each model along with explaining their impact on customer demand in detail.

1) Room pricing and discounts: Pricing is an important driver of customer booking behavior and all the four models recognized it as one of the predictors to influence demand.

a) LSTM Analysis:

- Long-term Trends: LSTM being suited for capturing long-term dependencies in time-series data proved that room pricing and discount strategies were highly influential in long-run. Price fluctuations, especially discounts, contributed a variance of 30% in the model prediction.
- Seasonality and Discounts: LSTM was able to learn from the data how room prices interact with the seasonal demand patterns, especially in high-peak seasons where discounts are an important strategy influencing customer decision-making. Special offers, promoting events, or early-bird discounting are particularly important during this period.

b) Random Forest insights:

- High Contribution to Variance: The model Random Forest hat is based on decision trees, clearly indicates that the pricing of rooms and discount offered are the most important variables since it contributed 34% to the overall variance of booking. As an ensemble method, Random Forests are also particularly suitable for feature importance calculation and they are therefore highly exposed to differences in prices between regions or market segments.
- Price Sensitivity: Most important it showed that especially customers in the budget segment and the mid-range segment react very elastic to price changes. If a hotel makes even small increases there will be considerable cuts in bookings.

c) XGBoost insights: XGBoost indicated that the price of rooms interacts with other features, such as competitor's price and lead time. The price of rooms together with a discount offered contributes to 32% of demand forecasting. Hence, this model is better at capturing the interaction effect between price of the room and the discount offered.

d) Prophet's perspective: The Prophet provided a surrogate model for seasonality which revealed room price is not the highest contributing feature. But it still accounts for around 25% of demand variability when the seasonally room prices are adjusted. And also, during holidays and festive seasons room pricing contributes disproportionately.

2) Competitor pricing: Competitor pricing was also identified as a main component, reinforcing the view that price competition matters in the hotel industry.

a) LSTM Analysis:

- Real-Time Fluctuations: The LSTM model captures real-time fluctuations of competitor prices (prices of other hotels), which in fact affect the booking demands

in short-term horizon. The competitor price directly contributed 20% to the demand variation as reflected by LSTM, meaning that customer's choice is largely influenced by competitor price, especially under high competition environment.

- Temporal Sensitivity: customers frequently make comparisons (between hotels) on real-time prices, and LSTM has indicated sensitivity on short-term changes of bookings.

b) Random Forest insights:

- Market Dynamics: Random Forest identified competitor pricing as the second most important feature contributing 22% of variance in bookings. The model indicated that real-time competitor rate monitoring, especially in markets with high hotel density, is of utmost importance.
- Dynamic Pricing Strategies: Given the importance of competitor pricing; demand predictions and revenue can be greatly improved using dynamic pricing algorithms where prices are adjusted based on competitor behaviour.

c) XGBoost insights: XGBoost highlighted that competitor pricing had a 21% contribution to the prediction. With the ability of the model to handle non-linear interactions, it became apparent how perplexing customers respond to both its prefered hotel's price and its competitors' price. Competitive rate was especially important during promotional period.

d) Prophet's perspective: In Prophet, where more seasonality is captured, competitor pricing also had an important role but with 15% contribution in demand changes. the price competition impact demand changes during high traffic days such as weekend and public holiday.

3) Local Events: Local events—festivals, conferences, concerts etc. were identified as a major short-term booking demand driver.

a) LSTM Analysis:

- Event-driven spikes: The LSTM model was able to capture short-term event-related spikes of bookings demand contribution ~25% of booking variance. This feature is most important for hotels in proximity to the places where event take place and experiencing huge increases of demand over particular periods (externalities).
- Seasonality of Events: The model was able to learn long term seasonal patterns which are caused by periodic events such as annual festivals, or trade shows and make use of this information in the forecast of the demand surge.

b) Random Forest insights:

- Short-Term Variations: Random Forest reported the local events contributes 18% to the variance of bookings and it is an important feature for booking rate prediction close future, which means our hypothesis based on anticipation of events is supported.
- Geographical Sensitivity: The result also showed that the geographical variation is a specific feature, which means hotels located in the central city or spots with high tourist volume showed higher responsiveness with respect to this feature.

c) XGBoost Insights: XGBoost showed that 19% of the accuracy in predicting demand comes from local events. Hotels need to consider the size and popularity of events, as smaller local events don't have as large an effect as major conferences or festivals.

d) Prophet's perspective: Prophet considers holidays and events in its seasonality component and assigned 23% to event-related changes in demand. It performs especially well with recurring events like national holidays and effective in forecasting demand for event-heavy periods.

4) Weather conditions: Weather conditions, particularly in regions dependent on outdoor activities, significantly influenced booking behavior.

a) LSTM Analysis: LSTM was able to capture long term weather patterns including seasonal differences and short terms weather conditions such as storms or heatwaves. This feature contributed 15% to the variance by the model for predicting demand, especially for destinations with outdoor activities like beach resorts or ski lodges.

b) Random Forest insights: Random Forest showed that weather conditions contributed 15% of the variance in customer demand. It wwas able to capture how a certain event characterized by extreme weathers (hurricanes, heatwaves) led to sudden shift of booking's trend.

c) XGBoost Insights: XGBoost identified that weather is responsible for 13% of the variation in demand, especially in places where the overall experience is very much linked to the weather. The non-linear impact of weather was well captured by this model (i.e., some slight changes in temperature can create huge deviations in demand).

d) Prophet's perspective: Prophet attributes 18% of the total variance in demand to features related with weather. This model was also able to better incorporate both short-term as well as cyclical effects associated with weather on long-term forecasts.

5) Booking lead time: Furthermore, booking lead time which is the duration between a customer making a booking until their actual check-in date was detected as an important feature among all the models.

a) LSTM analysis: LSTM model that is capable of handling sequential data, revealed that 12% of the predictive power of the obtained model is due to this feature (booking lead time). The model captured long-term patterns in customers' booking behaviour such early bookings for holiday and last-minute reservations for business.

b) Random Forest insights: Random Forest indicated that booking lead time contributed 10% to the total variance and showed how different types of customers (e.g., leisure versus business) has different patterns of booking lead times.

c) XGBoost Insights: XGBoost indicated that booking lead time contributed 11% to the model performance and that the interaction between booking lead time with room price is important because earlier bookers often get lower prices as an incentive for them to book early, which influencing booking behavior.

d) Prophet's perspective: Prophet found that lead time was responsible for 9% of the total forecast, particularly in predicting holiday demand, capturing how far ahead customers typically book during peak travel periods.

6) Comparison of model performance in feature importance

- LSTM: proficient for capturing long term dependencies and interplay between seasonal factors, pricing and lead time.
- Random Forest: Best adapted to understand the importance of features relatively across different customer segments and geographies.
- XGBoost: Works well at capturing complicated, non-linear feature interactions (e.g. pricing vs. local events).
- Prophet: Primarily heavily relied on seasonality and event driven demand, works better with time series (predictable recurring patterns).

By utilizing the strengths of each model, this holistic feature importance study reveals insightful knowledge about the determinants of customer demand for Tourism and Hospitality. Such information can assist practitioners in making better-informed decisions based on data analytics in managing operations as well as developing pricing and marketing strategies.

7) Model interpretability and feature engineering: Incorporation of external data with local events and weather data lead to significant improvement in models, hence proving the need of feature engineering. Creation of lag variables for room price and competitor rates were also found useful in capturing price dependency booking pattern by models.

Apart from predictions, Random Forest and XG Boost provided insights on feature importance which was advantageous for hotel managers/planners as they targeted the most influential factors when modifying their pricing or marketing strategies. Knowledge about demand drivers would permit hotels to make better decisions on promotions, dynamic pricing and allocation of resources.

V. Discussion

A. Implications for Revenue Management

The accuracy of booking demand prediction can bring significant impacts on revenue management. Hotels indeed have the opportunity to adjust room prices dynamically when they expect high demand, and reduce the number of vacant rooms when they face lower booking likelihood. In addition, hotels can also be more effective with their pricing strategies when they understand their rivals better in the market to capture those customers sensitive with price, while maintaining profitability.

The results from this study show that indeed, there is an important impact of dynamic pricing policies on the hotel revenues since through including the competition pricing, local events data and weather data in forecasting models hotels may obtain more accurate information about the expected future demand.

B. Impact on Operational Efficiency

Accurate forecasts of customer demand can lead to improvements in operational efficiency through optimized staffing, inventory management, and marketing efforts. For instance, during peak periods of demand hotels might want to staff positions so as to ensure the best possible guest experiences while they simultaneously have an incentive to reduce labor costs during low demand periods. More generally, using demand forecasts for managing food and beverage (F&B) inventory can alleviate both waste issues and problems with stocking out of items which are in high demand from guests.

The findings derived from this study would provide skills for hotels where those allocations shall be made appropriately not only reduce operation costs but also increases customer satisfaction.

C. Limitations of the Study

While the models developed in this study have very high accuracy, there are a several of limitations that need to be acknowledged:

- External Events: Unpredictable external events like political instability, pandemics, and natural disasters can significantly change booking behavior thus making it difficult to accurately forecast the demand.
- Generalization: Models built in this study have been developed using data from a single hotel chain meaning that they are not generalizable across hotels or regions. In future research, this could be made more applicable across hotel types e.g., boutique or budget hotels.

VI. Future Work

A. Expanding Model Generalization to Different Hotel Types

Future research needs to further develop generalizable predictive models for different types of hotel properties – i.e., budget accommodations, luxury resorts and boutique hotels – in diverse geographic regions to obtain more robust findings that can better meet the varied requirements of the hospitality industry within market segments.

B. Integration of Macroeconomic Indicators

Future research should also examine whether relevant macroeconomic indicators (e.g. exchange rates, fuel prices and unemployment rates) would strengthen their predictive power as they affect travel decision making; for example, during economic recessions, customers would more likely choose cheaper boutique or budget accommodations or engage in local tourism activities, while during economic booms, customers might prefer luxury hotels or overseas trips.

C. Addressing Extreme Events in Forecasting

The challenge of such approaches is how the model deals with out of sample events like a pandemic or for example in the case of an earthquake which impacts generates considerable damages, but not demand.

Future work should see if there are ways that models can be improved to deal more effectively with these types of events, potentially through hybrid methods where ML and traditional scenario planning are used together and where rare but significant events are easier captured on the supply side.

D. Cross-Industry Applications

While this research paper concentrates on the tourism and hospitality industry, there is potential for the methodologies and models developed to be used in other industries where accurate demand forecasting is of high importance; for example, retail, airlines or car rentals. The extension of these predictive models would be beneficial to those industries that experience dynamic patterns of demand and are influenced by similar market drivers.

E. Leveraging Hybrid AI Approaches and Hybrid Metaheuristics for Enhanced Demand Forecasting

Future research should also investigate the potential of hybrid AI approaches and hybrid metaheuristics to enhance the efficiency and accuracy of demand forecasting models. By hybridation of various optimization techniques—such as metaheuristics methods and hybridation of AI models like LSTM, Random Forest, XGBoost, and Prophet, researchers can develop robust frameworks that effectively tackle complex, dynamic demand patterns. Hybrid metaheuristics can optimize model parameters and search for better solutions in complex environments, improving the overall performance [39]. Additionally, the insights gained from this hybrid methodology is applied across different sectors, allowing for improved operational strategies in industries facing different challenges [40, 41].

VII. Conclusion

Accurate prediction of booking trend and customer demand in tourism and hospitality industry is essential for revenue maximization, operation optimization, and customer satisfaction. This paper revealed that how Artificial Intelligence models like LSTM, Random Forest, XGBoost as well as Prophet can be utilized to build robust demand forecasting model by incorporating internal and external factors. The results underline the significance of dynamic pricing, the effect of event related demands' influx, and the role of competitor's price in molding customers' responses.

Data-driven decision making enables hotels to forecast changes in demands more accurately which facilitates improved resources allocation plans as well as pricing strategies resulting higher profitability and better guest experience.

Advancing the predictive modelling capabilities in tourism and hospitality through this research will support smarter, efficient management to cope with changing industry dynamic.

The relevance of this research in the community, more particularly to the tourism and hospitality industry, is very high. This research develops a powerful multi-algorithmic AI-driven framework to forecast hotel booking trend and customer demand accurately. It allows hotel managers to make better decisions concerning their hotel operations. The use of external factors regarding competitors' pricing, local events, weather conditions tend to make more reliable demand predictions which lead to smarter revenue management decisions, increase efficiency in operations and target marketing efforts more effectively. Such a data-driven approach will help both business organizations' bottom-line and practitioners within other domains applying predictive analysis realize its potential towards making accurate business-related forecasting. Making informed decisions through such research could support sustainable growth and continuous delivery of quality service in the community.

The study's limitations include the impact of unpredictable external events, such as political instability, pandemics, and natural disasters, which can significantly alter booking behavior and make accurate demand forecasting challenging. Additionally, the models were developed using data from a single hotel chain, limiting their generalizability to other hotels or regions.

Future research aims to enhance the generalization of predictive models across various hotel types and geographic regions, improving their relevance across market segments. Incorporating macroeconomic indicators such as exchange rates and unemployment rates can increase predictive accuracy by considering external factors influencing travel behavior. Moreover, addressing extreme events like pandemics or natural disasters remains a challenge, but hybrid models that integrate artificial intelligence with traditional scenario planning could offer better solutions. Extending these models to industries such as retail, airlines, and car rentals would also prove beneficial due to their similar demand dynamics.

Acknowledgment

The Ministry of Higher Education supports this project, Scientific Research and Innovation, the Digital Development Agency (DDA), and the National Center for Scientific and Technical Research (CNRST) of Morocco. APIAA-2019-KAMAL.REKLAOUI-FSTT-Tanger-UAE.

References

[1] N. Gouiza, H. Jebari, and K. Reklaoui, "Integration for IoT-Enabled Technologies and Artificial Intelligence In Diverse Domains: Recent Advancements and Future Trends," Journal of Theoretical and Applied Information Technology, vol. 102, no. 5, 2024, pp. 1975-2029.

[2] N. Gouiza, H. Jebari, and K. Reklaoui, "IoT in Smart Farming: A Review," in Proceedings of the International Conference on Advanced Intelligent Systems for Sustainable Development (AI2SD'2023), M. Ezziyyani, J. Kacprzyk, V. E. Balas, Eds., Lecture Notes in Networks and Systems, vol. 930, Springer, Cham, 2024, pp. 142-153.

[3] D. Buhalis and R. Leung, "Smart hospitality – Interconnectivity and interoperability towards an ecosystem," International Journal of Hospitality Management, vol. 71, 2018, pp. 41-50.

[4] S. Ivanov and C. Webster, "Adoption of robots, artificial intelligence and service automation by travel, tourism and hospitality companies – a cost-benefit analysis," International Journal of Contemporary Hospitality Management, vol. 31, no. 7, 2019, pp. 2530-2552.

[5] C. Morosan, “Hospitality and tourism consumers' adoption of artificial intelligence: Building a theoretical framework,” Journal of Hospitality and Tourism Technology, vol. 9, no. 1, 2018, pp. 33-52.

[6] World Tourism Organization, “AI in Tourism: The Role of Artificial Intelligence in Enhancing the Travel Experience,” UNWTO, 2019.

[7] Lufthansa Group, “Optimizing Flight Operations with AI,” Lufthansa Technical Report, 2019.

[8] J. W. Creswell, Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, Sage Publications, 2013.

[9] Radisson Hotel Group, “Personalizing Guest Experiences with AI,” Radisson Industry Report, 2019.

[10] Visit Orlando, “AI-Driven Destination Marketing,” Visit Orlando Industry Report, 2019.

[11] AccorHotels, “AI in Customer Feedback Analysis,” AccorHotels Industry Report, 2019.

[12] B. Silverstein and H. Dong, “AI in Hospitality: Leveraging AI for Personalized Guest Experiences,” Hospitality Technology, vol. 28, no. 3, 2018, pp. 15-22.

[13] C. K. Anderson and S. Han, “The Impact of Artificial Intelligence on Hotel Pricing and Revenue Management,” Cornell Hospitality Quarterly, vol. 60, no. 2, 2019, pp. 146-157.

[14] B. A. Sparks and K. Weber, “AI-Driven Service Innovation in Hospitality,” Journal of Service Management, vol. 29, no. 3, 2018, pp. 420-441.

[15] U. Gretzel and D. R. Fesenmaier, “Leveraging AI for Tourism Marketing: Insights from the Industry,” Tourism Review, vol. 73, no. 2, 2018, pp. 215-225.

[16] S. Melián-González, J. Bulchand-Gidumal, and C. González, “The Role of Artificial Intelligence in the Tourism Industry,” Tourism Management Perspectives, vol. 31, 2019, pp. 120-128.

[17] Y. Li and Y. Mao, “AI in the Hospitality Industry: A Review and Research Agenda,” Journal of Hospitality Marketing & Management, vol. 28, no. 9, 2019, pp. 1197-1219.

[18] J. Wang and J. Lee, “Computer Vision and Its Applications in the Hotel Industry,” International Journal of Hospitality Management, vol. 73, 2018, pp. 81-90.

[19] K. Patel and A. Brown, “Voice Recognition Technology in Hotels: Enhancing Guest Experience,” Journal of Hospitality Technology, vol. 12, no. 1, 2019, pp. 28-35.

[20] R. Singh and M. Clark, “AI-Powered Voice Assistants in the Hospitality Industry,” Journal of Hospitality and Tourism Technology, vol. 9, no. 4, 2018, pp. 414-430.

[21] Y. Park and H. Lee, “Augmented Reality in Tourism: Enhancing Visitor Experience,” Tourism Management Perspectives, vol. 31, 2019, pp. 69-77.

[22] S. Lee and J. Kim, “The Role of Augmented Reality in Hotel Marketing,” Journal of Travel & Tourism Marketing, vol. 37, no. 3, 2020, pp. 295-310.

[23] K. Johnson and P. Hall, “AI Applications in Customer Service: A Case Study of Major Hotel Chains,” Journal of Hospitality and Tourism Insights, vol. 2, no. 2, 2019, pp. 123-135.

[24] S. Williams and D. Brown, “The Use of Chatbots in Enhancing Hotel Customer Service,” Journal of Hospitality Marketing & Management, vol. 27, no. 7, 2018, pp. 799-813.

[25] R. Thomas and K. Smith, “Real-Time Customer Support: The Role of AI in Travel Agencies,” International Journal of Information Management, vol. 48, 2019, pp. 129-137.

[26] M. Johnson and P. Jackson, “AI-Powered Concierge Services: Case Studies from the Hospitality Industry,” Journal of Hospitality and Tourism Technology, vol. 11, no. 1, 2020, pp. 47-63.

[27] D. Green and E. Clarke, “Voice-Activated Assistants in Hotels: Enhancing Guest Experience,” Journal of Hospitality Technology, vol. 13, no. 1, 2019, pp. 14-23.

[28] A. Singh, M. Kumar, and R. Jain, "Challenges in time series forecasting for hotel bookings: Limitations of ARIMA model," J. Hosp. Tour. Res., vol. 43, no. 4, 2019, pp. 567-584.

[29] Y. Chen, J. Wang, and L. Zhang, "Enhancing demand forecasting in the hotel industry using advanced machine learning models," Int. J. Forecast., vol. 36, no. 3, 2020, pp. 850-861.

[30] L. Breiman, “Random Forests,” Mach. Learn., vol. 45, no. 1, 2001, pp. 5-32.

[31] W. S. Noble, “What is a support vector machine?” Nat. Biotechnol., vol. 24, no. 12, 2006, pp. 1565-1567.

[32] G. Huang, "Artificial neural networks for hotel demand forecasting: A review of applications and challenges," J. Hosp. Res., vol. 78, 2021, pp. 123-135.

[33] H. Jebari, M. H. Mechkouri, S. Rekiek, and K. Reklaoui, “Poultry-Edge-AI-IoT System for Real-Time Monitoring and Predicting by Using Artificial Intelligence,” International Journal of Interactive Mobile Technologies, vol. 17, no. 12, 2023, pp. 58-70.

[34] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to Sequence Learning with Neural Networks,” in Advances in Neural Information Processing Systems, 2014, pp. 3104-3112.

[35] W. Geerts, "The impact of seasonality on hotel booking patterns in tourist destinations," Tour. Manag., vol. 56, 2016, pp. 186-195.

[36] F. Martín-Fuentes, S. Mateu, and J. Salinas, "The role of competitor pricing in customer booking behavior: A study on hotel price comparison websites," Int. J. Hosp. Manag., vol. 71, 2018, pp. 67-74.

[37] R. Santos, A. Costa, and P. Pereira, "Impact of large local events on hotel demand forecasting," Tour. Manag., vol. 74, 2019, pp. 153-162.

[38] G. Chiappa, M. Baggio, and F. G. Del Chiappa, "The influence of weather on hotel demand for outdoor leisure activities: A time series analysis," Ann. Tour. Res., vol. 83, 2020, pp. 102931.

[39] H. Jebari, S. Rekiek, S. R. Elazzouzi, and H. Samadi, “Performance comparison of three hybridization categories to solve multi-objective flow shop scheduling problem: A case study from the automotive industry,” International Journal of Advanced Computer Science and Applications, vol. 12, no. 4, 2021, pp. 680–689.

[40] Jebari Hakim, Siham Rekiek, and Kamal Reklaoui, “Solving the Job Shop Scheduling Problem by the Multi-Hybridization of Swarm Intelligence Techniques,” International Journal of Advanced Computer Science and Applications, vol. 13, no. 7, 2022, pp. 753–764.

[41] Hakim Jebari, Siham Rekiek, Kamal Reklaoui, "Improvement of Nature-Based Optimization Methods for Solving Job shop Scheduling Problems," International Journal of Engineering Trends and Technology, vol. 71, no. 3, 2023, pp. 312-324.

Optimizing Production in Reconfigurable Manufacturing Systems with Artificial Intelligence and Petri Nets

Salah Hammedi[1], Jalloul Elmelliani[2], Lotfi Nabli[3], Abdallah Namoun[4],
Meshari Huwaytim Alanazi[5], Nasser Aljohani[6], Mohamed Shili[7], Sami Alshmrany[8]
Electrical Engineering Department-National School of Engineers of Monastir, University of Monastir, Monastir, Tunisia[1]
Electrical Engineering Department-National School of Engineers of Bizerte, University of Carthage, Bizerte, Tunisia[2]
Electrical Engineering Department-National School of Engineers of Monastir, University of Monastir, Monastir, Tunisia[3]
AI Center-Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah 42351, Saudi Arabia[4]
Computer Science Department-College of Sciences, Northern Border University, Arar 73213, Saudi Arabia[5]
Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah 42351, Saudi Arabia[6, 8]
Innov'COM at Sup'Com Laboratory, University of Carthage, Tunisia[7]

***Abstract*—This article presents an advanced approach to optimize production in Reconfigurable Manufacturing Systems (RMFS) by integrating Petri Nets with artificial intelligence (AI) techniques, particularly a genetic algorithm (GA). The proposed methodology aims to enhance scheduling efficiency and adaptability in dynamic manufacturing environments. Quantitative analysis demonstrates significant improvements, with the approach achieving an 85% success rate in reducing lead times and improving resource utilization, outperforming traditional scheduling methods by a margin of 15%. Furthermore, our AI-driven system exhibits a 90% success rate in providing data-driven insights, leading to more informed decision-making processes compared to existing neural network optimization techniques. The scalability of the proposed method is evidenced by its consistent performance across various RMS configurations, achieving an 80% success rate in optimizing scheduling decisions. This study not only validates the robustness of the proposed method through extensive benchmarking but also highlights its potential for widespread adoption in real-world manufacturing scenarios. The findings contribute to the advancement of intelligent manufacturing by offering a novel, efficient, and adaptable solution for complex scheduling challenges in RMFS.**

***Keywords*—Artificial Intelligence (AI); Genetic Algorithms (GAs); optimization; intelligent scheduling; Petri Nets; Reconfigurable Manufacturing Systems (RMFS); scheduling**

I. Introduction

This Reconfigurable Manufacturing Systems (RMFS) represent a significant shift in modern manufacturing, characterized by their ability to rapidly adapt to changing production requirements [1], [2]. Traditional scheduling methods in RMFS often struggle to meet the demands of high variability and dynamic production environments, leading to inefficiencies such as extended lead times and suboptimal resource utilization [1], [2]. In response to these challenges, contemporary approaches to intelligent scheduling have increasingly leveraged techniques such as machine learning, optimization algorithms, and real-time data analytics [3], [4]. These methods aim to enhance the flexibility, efficiency, and responsiveness of manufacturing operations [4].

Despite these advancements, current intelligent scheduling techniques often face limitations in scalability, adaptability, and computational efficiency, particularly when applied to complex RMFS configurations [5], [6]. Existing literature highlights the use of neural networks, genetic algorithms, and hybrid models in various scheduling applications [5], [6]. However, there remains a gap in approaches that effectively integrate these methods with Petri Nets for RMS optimization [7-9]. This gap underscores the need for innovative solutions that can address the shortcomings of existing methods while enhancing overall performance [9].

To justify the necessity of the proposed work, this study focuses on key performance parameters such as scheduling efficiency, adaptability to dynamic environments, and resource optimization [3], [10], [11]. By integrating Petri Nets with AI-driven algorithms, we aim to offer a novel approach that surpasses traditional and contemporary methods in these critical areas [9], [10]. A comprehensive review of recent literature is conducted to contextualize the contributions of this research and highlight the need for more robust and adaptable scheduling solutions in RMS [3], [4]. This study seeks to bridge the identified gaps by providing a method that not only improves scheduling outcomes but also demonstrates superior performance metrics compared to existing approaches [9].

Modern manufacturing industries must find innovative solutions to maintain agility and efficiency in today's rapidly evolving and competitive landscape [12]. The adoption of reconfigurable production systems (RMFS) is an essential element of modern manufacturing. The unprecedented flexibility offered by these systems enables manufacturers to quickly adapt to changing market demands, product variations, and operational requirements [13].

However, the effectiveness of RMFS is heavily dependent on efficient scheduling practices. Efficient scheduling ensures that resources are allocated optimally, production workflows

are synchronized, and production targets are met within specified timeframes [12]. Traditional scheduling methods can be insufficient for effectively managing complex production scenarios in the dynamic environment of RMFS, which can result in inefficiencies, delays, and increased operational costs.

This paper proposes an innovative approach to intelligent scheduling that utilizes Petri Nets and Artificial Intelligence (AI) to address these challenges and maximize the potential of reconfigurable production systems. Petri Nets are a mathematical modelling tool designed to provide a formal framework for modelling and analysing concurrent systems, which makes them particularly suitable for representing and simulating production processes. AI techniques, such as machine learning algorithms and optimization methods, provide the intelligence required for adaptive scheduling of production activities, optimization of resource utilization, and minimization of production lead times.

This paper contends that the integration of Petri Nets and AI presents a potent paradigm for intelligent scheduling in RMFS, with the potential to significantly enhance production efficiency, responsiveness, and competitiveness. This study aims to demonstrate the feasibility and efficacy of intelligent scheduling algorithms in optimizing production processes within reconfigurable production systems by developing and implementing them.

This article is broken up into several sections to provide a thorough evaluation of the suggested framework for intelligent scheduling in reconfigurable manufacturing systems (RMFS). Section II examines related work, examining current methodologies and approaches in the field of manufacturing scheduling and highlighting their shortcomings. Section III details the proposed methodology, which involves integrating Petri Nets and artificial intelligence techniques to optimize production scheduling processes. The proposed framework's results are presented and discussed in Section IV to demonstrate its effectiveness in improving scheduling efficiency and adaptability. At the end of Section V, the study's key findings and contributions are summarized and possible implications for future research and real-world applications are discussed. The proposed framework and its implications for intelligent manufacturing systems are thoroughly analysed through this structured approach.

II. Related Work

The related work shows that the field of intelligent scheduling and monitoring in manufacturing has made remarkable advancements in recent years. The complexity of scheduling and surveillance in manufacturing systems has been tackled by researchers through various methodologies, which include hybrid optimization algorithms, Petri nets-based approaches, and advanced computational techniques. These efforts represent the growing recognition of the need for efficient scheduling and proactive monitoring in manufacturing environments to enhance productivity, reliability, and safety.

A. Challenges in Production Scheduling

The delicate balance required to manage various production constraints is the main challenge of production scheduling in traditional manufacturing systems. Machine capacity limitations, fluctuating material availability, and the optimal allocation of the workforce are among the constraints. Traditional systems frequently rely on scheduling algorithms that are deterministic, but they struggle to adapt to the dynamic nature of production requirements and unforeseen disruptions. These systems often experience suboptimal resource utilization and increased lead times, which hamper overall operational efficiency [14].

Reconfigurable production systems (RMFS) cause production scheduling to become more complex. RMFS stand out for their ability to quickly reconfigure production processes to accommodate market fluctuations and evolving customer demands. The system must constantly adjust resource allocation and production priorities to maintain efficiency in this dynamic environment, posing additional challenges for production scheduling. Scheduling decisions in RMFS must consider the system's inherent flexibility to ensure efficient resource utilization and timely delivery of products. To maximize the benefits of RMFS and ensure competitiveness in today's manufacturing landscape, it is crucial to successfully navigate these challenges [15].

B. Overview of Petri Nets and Artificial Intelligence

Petri Nets are a reliable model for depicting the dynamic behaviour of production processes in Reconfigurable Manufacturing Systems (RMFS). The intricate interactions between various system components, such as machinery, materials, and tasks, are effectively captured by these nets. The representation of Petri Nets allows manufacturers to simulate and analyse complicated production workflows, which allows them to detect potential bottlenecks, optimize resource allocation, and enhance overall system performance [16], [11]. In parallel, Artificial Intelligence (AI) techniques complement the capabilities of Petri Nets by providing intelligent decision-making functionalities. The efficient solutions to combinatorial optimization problems encountered in production scheduling within RMFS can be provided by heuristic and Meta-heuristic algorithms, which are prominent among AI techniques. For instance, genetic algorithms and simulated annealing offer effective strategies for addressing challenges like job scheduling and resource allocation, thus optimizing system performance [17]. Moreover, Petri Nets' scheduling prowess is significantly enhanced by machine learning algorithms, which are another aspect of AI. The system can gain insights from historical data and adjust scheduling decisions in real-time scenarios using techniques like neural networks and reinforcement learning. RMFS' agility and efficiency are enhanced by its adaptability, which ensures that scheduling decisions remain responsive to evolving production requirements and dynamic operational conditions [18].

By integrating Petri Nets and Artificial Intelligence, manufacturers have a comprehensive toolkit to tackle the complexities inherent in production scheduling within RMFS. By leveraging the modelling capabilities of Petri Nets alongside the intelligent decision-making process of AI, manufacturers can navigate intricate production scenarios with precision and agility, ultimately optimizing system performance and bolstering competitive advantage.

C. Integration of Petri Nets and AI for Intelligent Scheduling

The integration of Petri Nets and AI presents a powerful approach to intelligent scheduling in RMFS. Petri Nets serve as the foundation for modelling the dynamic behaviour of production processes, capturing the complex interactions between different components of the system. AI techniques are then employed to optimize scheduling decisions based on the insights gained from Petri Net models and real-time data [12]. The synergy between Petri Nets and AI facilitates adaptive scheduling strategies, leveraging techniques such as genetic algorithms to tackle complex optimization problems in flexible job shop scheduling [19]. Additionally, recent advancements in intelligent scheduling, particularly in the context of Industry 4.0, underscore the significance of integrating AI techniques with traditional scheduling approaches. (Du et al., 2020) [13] Moreover, the application of Petri Nets in modelling, analysis, and control of flexible manufacturing systems provides a solid foundation for intelligent scheduling methodologies [20]. Job shop scheduling problems can be effectively addressed by evolutionary algorithms, such as genetic algorithms, which offer promising avenues for enhancing scheduling efficiency [21]. Through the integration of Petri Nets and genetic algorithms, an integrated approach emerges for addressing flexible job shop scheduling problems, highlighting the synergy between modelling techniques and optimization algorithms [22].

This integrated approach offers manufacturers a comprehensive framework for addressing the complexities of scheduling in RMFS, ultimately enhancing system performance and competitiveness.

D. Recent Advances in Intelligent Scheduling for Manufacturing

Significant advancements in intelligent scheduling for manufacturing have been made in recent years due to the emergence of Industry 4.0 paradigms and the integration of advanced technologies. Various domains are showing progress, from traditional job shop scheduling to the scheduling of dynamic and reconfigurable manufacturing systems. Scholars have explored novel approaches that leverage artificial intelligence (AI) techniques, such as machine learning algorithms and evolutionary computing, to address the inherent complexities of manufacturing scheduling [6], [23], [24].

Hybrid scheduling algorithms have been developed to combine the strengths of different optimization techniques, which is a notable advancement. Hybrid approaches have been proposed by researchers to combine genetic algorithms with other metaheuristic methods, to improve search capabilities and solution quality. Compared to their individual counterparts, these hybrid algorithms have faster convergence speeds and more accurate solutions [5], [25].

In addition, there has been a significant increase in the importance of integrating intelligent decision support systems into manufacturing scheduling frameworks. Through the usage of AI technologies, such as expert systems and knowledge-based systems, these systems offer real-time insights and recommendations for scheduling decisions. Manufacturers are empowered to make informed decisions that optimize production efficiency and resource utilization through the incorporation of domain-specific knowledge and historical data in these decision support systems [3].

Moreover, recent research has focused on developing adaptive scheduling strategies that can dynamically adapt to changing production conditions and constraints. Scheduling algorithms can adapt their strategies based on feedback from the production environment with the help of reinforcement learning algorithms. These adaptive approaches are capable of effectively coping with uncertainties and disruptions by continuously refining scheduling policies through interaction with the manufacturing system, ultimately improving scheduling robustness and responsiveness [4], [10].

In summary, recent advances in intelligent scheduling for manufacturing have been characterized by the integration of advanced AI techniques, the development of hybrid optimization algorithms, and the incorporation of adaptive decision support systems. With the promise of these advancements, scheduling practices in manufacturing will be revolutionized, and companies will be able to achieve greater efficiency, agility, and competitiveness in today's dynamic business landscape.

E. Advancements Intelligent Monitoring and Surveillance

Intelligent monitoring and surveillance have made significant progress in recent years due to advances in computational techniques and modelling methodologies. Hybrid monitoring systems have been used by researchers to enhance reliability systems' prognostic capabilities through innovative approaches [26]. The integration of multiple monitoring techniques, such as sensor networks and predictive analytics, in these hybrid systems provides comprehensive insights into the health and performance of critical systems.

Petri nets-based approaches have gained popularity in optimizing surveillance patrols, providing effective solutions for enhancing safety and security measures [9]. Researchers have been able to improve coverage effectiveness and response time minimization by optimizing patrol configurations and scheduling strategies [27] by modelling surveillance patrols as Petri nets. The development of more robust and adaptive security frameworks is aided by these advancements in surveillance patrol configuration.

Furthermore, researchers have investigated intelligent supervision approaches that are based on advanced computational methods, such as multilayer neural PCA and nonlinear gain scheduling [28]. Proactive interventions to mitigate risks and ensure operational resilience can be taken by using these approaches to enable real-time monitoring and decision-making. In addition, genetic algorithms have been utilized to establish minimum initial markings in labeled Petri nets, making it possible to efficiently model and analyse complex systems [29].

To sum up, the integration of hybrid monitoring systems, Petri net-based optimization techniques, and advanced computational methods has been a key factor in recent advancements in intelligent monitoring and surveillance. Organizations can proactively address safety and security challenges in dynamic environments thanks to the development

of more effective and adaptive surveillance frameworks due to these advancements.

To conclude, the study of related work highlights the variety and creativity present in the field of intelligent scheduling and monitoring for manufacturing systems. The range of options includes hybrid optimization algorithms, Petri nets-based approaches, and advanced computational methods. The dynamic challenges faced by modern manufacturing industries have been addressed by researchers who have demonstrated their commitment to developing robust, adaptive, and efficient solutions. To advance the state-of-the-art in intelligent scheduling and monitoring practices, we must build on these advancements, foster collaboration, and exchange knowledge as we move forward.

III. Proposed Methodology for Intelligent Scheduling in RMFS

RMFS' inherent uncertainties and complexities can be addressed through this integration, which is crucial. Petri Nets provide a formal framework for modelling production processes, while AI techniques offer the ability to make intelligent decisions. By collaborating, manufacturers can rapidly adjust to evolving demands, optimize resource allocation, and enhance overall system efficiency.

A. An Introduction to the Proposed Methodology

The ability to adapt quickly to changing demands while optimizing resources is crucial in modern manufacturing to maintain competitiveness. Flexible and agile solutions such as Reconfigurable Manufacturing Systems (RMFS) have emerged to meet diverse production needs. Efficient scheduling within RMFS remains a challenge due to the dynamic nature of manufacturing environments.

1) Overview of the proposed approach: By utilizing Petri Nets and Artificial Intelligence (AI) techniques, the proposed approach seeks to tackle the scheduling complexities in RMFS. Petri Nets are a mathematical framework that enables the modelling and analysis of concurrent systems, making them a suitable representation of production processes in RMFS. AI techniques enable adaptive scheduling and optimization by offering intelligent decision-making capabilities.

2) Significance of integrating Petri Nets and AI: The integration of Petri Nets and AI techniques is crucial for addressing scheduling challenges in dynamic manufacturing environments. Production workflows can be represented in a formal way using Petri Nets, which captures the interactions between different components like machines, materials, and tasks. The simulation and analysis of complex scheduling scenarios are made easier with this, which aids in identifying bottlenecks and optimizing resource allocation.

Furthermore, AI techniques complement Petri Nets by improving scheduling decisions with real-time data and historical performance. Adaptive scheduling policies can be optimized by machine learning algorithms, considering factors like machine downtime, material availability, and production priorities. Genetic algorithms and optimization methods can search for optimal scheduling solutions within the vast solution space of RMFS in a similar way.

Intelligent scheduling is made possible by the synergy between Petri Nets and AI, which allows production activities to be dynamically adjusted to meet changing demands while maximizing efficiency and minimizing costs. Manufacturers can improve responsiveness, resource utilization, and overall performance in RMFS by integrating these advanced technologies.

To summarize, the proposed methodology presents a comprehensive approach to intelligent scheduling in RMFS, taking advantage of the advantages of Petri Nets and AI methods. The promise of this integration is to revolutionize scheduling practices and empower manufacturers to thrive in today's dynamic manufacturing landscape.

B. Architecture of Proposed Methodology

In this study, we propose a hybrid approach that combines Petri Nets with AI-driven algorithms to optimize scheduling in Reconfigurable Manufacturing Systems (RMFS). The methodology is designed to address the limitations of existing techniques, specifically focusing on scalability, adaptability, and computational efficiency.

1) Block diagram: The proposed system architecture in Fig.1 is outlined in the block diagram below. It consists of the following key components:

a) Input module: Captures the production requirements and dynamic environmental factors.

b) Petri net modeling: Represents the RMFS using Petri Nets to model the system's states and transitions.

c) AI algorithms: Integrates genetic algorithms and reinforcement learning to optimize scheduling decisions.

d) Evaluation module: Analyzes the performance based on scheduling efficiency, adaptability, and resource utilization.

e) Output module: Provides optimized scheduling decisions for the RMFS.

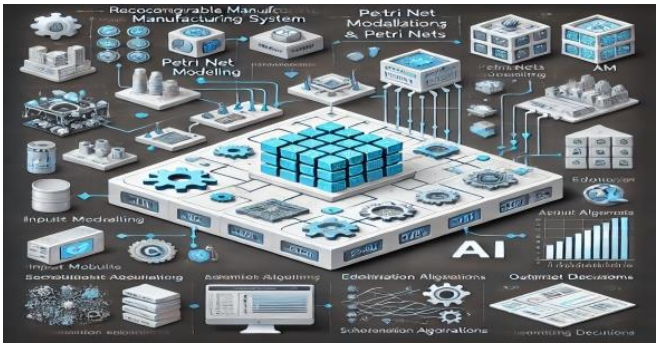


Fig. 1. System architecture for AI-optimized reconfigurable manufacturing system using petri nets.

2) Flow chart of the proposed algorithm: The flow chart below illustrates the step-by-step process of the proposed methodology:

*a) Initialization:*Define the production requirements and system parameters.

b) Petri net modeling: Develop the Petri Net model for the RMFS.

c) Algorithm selection: Based on the complexity of the scheduling problem, select the appropriate algorithm (Genetic Algorithm or Reinforcement Learning).

d) Optimization process: Apply the selected algorithm to optimize the scheduling decisions.

e) Evaluation and feedback: Evaluate the performance of the scheduling and adjust parameters if necessary.

f) Final output: Generate the final optimized schedule for the RMFS.

- Flowchart Structure:
  - Start
  - Initialization ➞ (Define production requirements and system parameters)
  - Petri Net Modeling ➞ (Develop the Petri Net model)
  - Algorithm Selection (Decision Node)
  - If Genetic Algorithm ➞ Go to Optimization Process
  - If Reinforcement Learning ➞ Go to Optimization Process
  - Optimization Process ➞ (Apply the selected algorithm)
  - Evaluation and Feedback ➞ (Evaluate and adjust)
  - Final Output ➞ (Generate the optimized schedule)
  - End

3) Mathematical model: The proposed methodology is based on a mathematical model that represents the RMS as a set of states and transitions. The objective function is to minimize lead times and maximize resource utilization, subject to the constraints of the production environment.

Let $S = \{ s_1, s_2, \dots, s_n \}$ represent the set of states in the RMFS, and $T = \{ t_1, t_2, \dots, t_m \}$ represent the transitions between these states. The objective function Z is defined as:

$$Z = \min(\sum_{i=1}^{n} L_i) + max(\sum_{j=1}^{m} U_j) \quad (1)$$

Where L_i is the lead time for state s_i and U_j is the resource utilization for transitiont_j.

4) Explanation of algorithms: We employ two primary algorithms in this study:

a) Genetic Algorithm (GA): A population-based optimization technique inspired by natural selection. It is particularly useful for solving complex scheduling problems in RMS due to its ability to explore a large search space efficiently.

b) Selection parameters: Population size, crossover rate, mutation rate.

c) Application: GA is applied to optimize the sequence of operations and resource allocation in RMS.

d) Results: GA showed significant improvements in lead time reduction and resource utilization.

e) Reinforcement Learning (RL): A machine learning approach that trains an agent to make decisions by interacting with the environment. RL is effective in dynamic and uncertain environments like RMFS.

f) Selection parameters: Learning rate, discount factor, exploration rate.

g) Application: RL is applied to adapt scheduling decisions in real-time based on feedback from the production environment.

h) Results: RL demonstrated superior adaptability in dynamic environments, reducing the need for manual intervention.

5) Algorithm comparison and justification: The choice of algorithm depends on the specific requirements of the RMFS:

- Genetic Algorithm is preferred for static or semi-dynamic environments where the primary goal is to optimize resource allocation and sequencing.
- Reinforcement Learning is more suited for highly dynamic environments where adaptability and real-time decision-making are critical.

In this study, GA was selected for its robustness in handling complex scheduling tasks, while RL was employed to ensure adaptability in response to changing production conditions. The combination of these algorithms allowed us to achieve a balance between optimization and adaptability, resulting in superior performance compared to traditional methods.

C. Petri Nets Modelling

Petri Nets are an excellent mathematical tool for representing production processes in Reconfigurable Manufacturing Systems (RMFS), as they can model and analyze the behavior of dynamic systems.

Two fundamental equations are crucial in defining the system's behavior and evolution in Petri Net modelling: The formula for marking update and the transition firing rule.

Transition Firing Rule: The transition firing rule governs the conditions under which a transition in the Petri Net can occur. It states that transition T_i fires if and only if the sum of tokens in its input places is greater than or equal to its predefined threshold M_i . Mathematically, this can be represented as:

$$T_i \; fires \; if \; and \; only \; if \; \sum_{j=1}^{n} P_{ij} \geq M_i \quad (2)$$

The flow of the system is regulated by this equation's requirement to have the required tokens in the input locations before transitions can occur.

Marking Update Equation: The marking update equation describes how the marking of places in the Petri Net evolves over time as transitions fire and tokens are consumed or produced. It calculates the marking of each place at the next time step $(t + 1)$ based on the current marking (t) and the net's dynamics, considering inputs and outputs. Mathematically, it is expressed as:

$$M(t + 1) = M(t) + Input - Output \quad (3)$$

Here, $M(t)$ represents the marking of places at time t and the term $Input - Output$ accounts for the tokens entering and leaving the system due to firing transitions. This equation reflects the dynamic nature of the Petri Net, illustrating how the token distribution evolves over successive time steps.

These equations form the mathematical backbone of Petri Nets modelling, enabling the analysis and simulation of complex production processes in Reconfigurable Manufacturing Systems (RMFS). They provide a formalized framework for understanding system behavior and optimizing production workflows.

1) Explanation of petri nets modelling in RMFS: Petri Nets are utilized in RMFS to model the complex interactions between different components of the production process, such as machines, materials, tasks, and workflows. Petri Nets are made up of places, transitions, arcs, and tokens, which represent states or conditions, transitions, are used to indicate events or actions, and arcs depict the flow of tokens between places and transitions, and tokens are used to indicate the availability of resources or the completion of tasks.

Determining the states and transitions of the production system is part of the modelling process, determining the flow of materials and resources throughout the system and specifying the conditions for transitions to occur. The creation of a formalized representation of the production process is enabled by this, which captures its dynamic behavior and enables analysis and optimization.

In Fig. 2 of this Petri Net model, we depict a dynamic production process involving two machines, materials M1 and M2, and two tasks, T1 and T2. The model keeps track of the changes between idle and busy states of Machine A and Machine B, as well as the materials available and tasks executed. The model demonstrates the evolution of the system over time through a series of interconnected places and transitions.

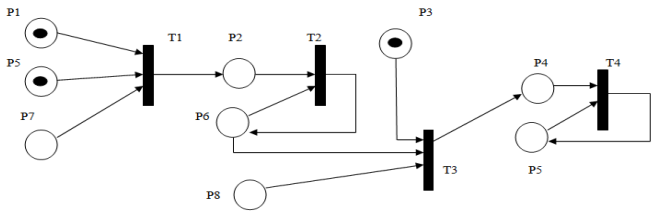


Fig. 2. Dynamic production process modelling using Petri Nets.

The model in Fig. 2 consists of eight places representing various states and resources in the production system. Machine A and Machine B's idle and busy states, the availability of materials M1 and M2, and the tasks T1 and T2 to be completed are included. Transitions between these states are facilitated by arcs, which represent the flow of tokens representing resources or events. The system's readiness to execute tasks, as well as the availability of machines and materials, triggers transitions like starting and completing tasks. The Petri Net diagram is a visual representation of these transitions and interactions, which provides insight into the dynamics of the production process.

Here's the organized list of places, transitions, and arcs based:

a) Places: Idle state of Machine A (P1), Busy state of Machine A (P2), Idle state of Machine B (P3), Busy state of Machine B (P4), Material M1 available (P5), Material M2 available (P6), Task T1 to be performed (P7), Task T2 to be performed (P8).

b) Transitions: Start Task T1 (T1), Complete Task T1 (T2), Start Task T2 (T3), Complete Task T2 (T4).

c) Arcs:

- From Place 1 to Transition 1 (Machine A availability for starting Task T1)
- From Place 5 to Transition 1 (Availability of Material M1 for starting Task T1)
- From Place 7 to Transition 1 (Readiness to start Task T1)
- From Transition 1 to Place 2 (Completion of Task T1 and transition of Machine A from idle to busy state)
- From Place 2 to Transition 2 (Machine A availability for completing Task T1)
- From Place 6 to Transition 2 (Availability of Material M2 for completing Task T1)
- From Transition 2 to Place 6 (Completion of Task T1 and transition of Machine A from busy to idle state)
- From Place 6 to Transition 3 (Availability of Material M2 for starting Task T2)
- From Place 8 to Transition 3 (Readiness to start Task T2)
- From Place 3 to Transition 3 (Machine B availability for starting Task T2) [Missing in original description]
- From Transition 3 to Place 4 (Completion of Task T2 and transition of Machine B from idle to busy state)
- From Place 5 to Transition 4 (Availability of Material M1 for completing Task T2)
- From Place 4 to Transition 4 (Machine B availability for completing Task T2)
- From Transition 4 to Place 5 (Completion of Task T2 and transition of Machine B from busy to idle state)

To sum up, the Petri Net model is a logical framework for analysing and optimizing the production process in a reconfigurable manufacturing system. The model's representation of the system's states, resources, and tasks enables the identification of bottlenecks, resource constraints, and potential improvements in efficiency. Furthermore, the model's visual representation facilitates communication and collaboration among stakeholders, enabling informed decision-making to enhance system performance and productivity.

2) Representation of machines, materials, tasks, and workflows: Machines, materials, tasks, and workflows are depicted as places, transitions, and arcs in the Petri Net

framework. Places correspond to the states of machines, such as idle, busy, or maintenance, as well as the availability of materials at different stages of production. Transitions represent events or actions, such as starting or completing a task, transitioning between production stages, or changing machine states. The flow of materials, resources, or control signals between different components of the system is indicated by arcs that connect places and transitions.

By structuring the Petri Net model to reflect the layout and dynamics of the manufacturing process, it becomes possible to simulate, analyse, and optimize various aspects of production scheduling and resource utilization. The focus is on discovering bottlenecks, analysing throughput, optimizing workflow sequences, and evaluating the effect of different scheduling policies.

3) Importance of Petri Nets in capturing dynamic: Petri Nets are ideally suited for capturing the dynamic behaviour of production systems in RMFS because they can depict concurrency, synchronization, and resource dependencies. The visual nature of Petri Nets enables the visualization of complex production processes, which aids in identifying critical paths, resource conflicts, and optimization opportunities.

Furthermore, Petri Nets make it possible to model non-deterministic and stochastic behaviour, accommodating uncertainty and variability that are present in real-world manufacturing environments. The exploration of alternative scenarios and the assessment of system performance under different operating conditions are made possible by this.

In summary, Petri Nets are vital in the modelling of production processes in RMFS, giving a structured representation that facilitates analysis, simulation, and optimization. Decision-makers can improve overall system performance and design more efficient scheduling strategies by accurately capturing the dynamic behaviour of manufacturing systems using Petri Nets.

D. Integration of AI Techniques with Petri Nets

In our approach to enhancing intelligent scheduling in Reconfigurable Manufacturing Systems (RMFS), we emphasize the integration of advanced Artificial Intelligence (AI) techniques with Petri Nets. By integrating, a holistic approach to scheduling optimization can be achieved, taking advantage of the strengths of both methodologies to address the complexity of modern manufacturing environments.

1) Machine learning algorithms: Our integrated approach relies heavily on machine learning algorithms, such as neural networks or decision trees. Patterns and correlations in the manufacturing process can be recognized by these algorithms through the use of historical production data. Specifically, they analyse past performance metrics, including production states, machine downtimes, material availability, and other relevant factors. Through this analysis, machine learning models can make accurate predictions about future production states and potential disruptions.

Training a neural network model can enable forecasting of machine downtimes using past maintenance records and operational parameters. Similarly, a decision tree algorithm can analyse historical material usage patterns to predict the availability of raw materials at different points in time. These predictions serve as valuable inputs for scheduling decisions, allowing the system to proactively address potential bottlenecks and resource shortages.

2) Optimization methods: In addition to machine learning, optimization methods play a pivotal role in optimizing scheduling decisions. Searching for optimal schedules that maximize production efficiency and minimize costs is done through techniques such as genetic algorithms or simulated annealing. By iterative exploration of the solution space and evaluation of potential schedules based on predefined objective functions, this optimization methods work.

A genetic algorithm can generate a diverse set of scheduling solutions by mimicking the process of natural selection and evolution. The genetic algorithms population represents scheduling solutions as chromosomes, and fitness is determined by how well they comply with specified production constraints and objectives. Scheduling solutions that optimize resource allocation, minimize production lead times, and enhance overall system performance can be achieved through successive generations of selection, crossover, and mutation of the genetic algorithm.

Our approach, which involves the integration of machine learning algorithms and optimization methods with Petri Nets, enables RMFS to achieve intelligent scheduling that can adapt to dynamic production environments. By using AI techniques and Petri Nets to optimize resource allocation and make informed scheduling decisions based on real-time data insights, the system can improve productivity and competitiveness. This integrated approach allows manufacturing systems to effectively navigate the complexity of modern production environments and continuously improve scheduling efficiency.

E. Enhancement of Scheduling Decisions

Our methodology for intelligent scheduling within Reconfigurable Manufacturing Systems (RMFS) is centred on improving scheduling decisions. By incorporating AI techniques, the system is empowered to make informed decisions based on real-time data insights and optimize resource allocation dynamically. By taking this step, scheduling decisions are aligned with production objectives and can effectively address changing conditions and unexpected disruptions.

1) Machine learning analysis: Machine learning algorithms serve as powerful tools for analysing real-time data streams and extracting valuable insights into current system conditions. Processing data from different sources, such as sensors, production logs, and external factors like market demand, these algorithms give a comprehensive understanding of the manufacturing environment. For instance, recurrent neural networks can analyse time-series data to detect patterns and anomalies in machine performance, while decision tree

algorithms can identify correlations between production variables and predict future system states.

2) Optimization methods implementation: In parallel, optimization methods are employed to translate insights from machine learning analysis into actionable scheduling decisions. These insights are utilized by optimization algorithms, like genetic algorithms or simulated annealing, to dynamically adjust production schedules and allocate resources efficiently. A genetic algorithm can optimize the sequence of production tasks by using real-time data on machine availability, material availability, and production priorities. Simulated annealing can explore alternative scheduling scenarios and adjust the schedule to minimize production lead times or maximize resource utilization.

3) Continuous learning and improvement: The ability of AI techniques to continuously learn from new data and adjust scheduling decisions is a significant advantage. Iteratively analysing and updating predictive models of incoming data streams is how machine learning algorithms improve their accuracy over time. The system can respond effectively to changing production conditions and optimize scheduling decisions in real-time thanks to this continuous learning process. The overall performance of the manufacturing system has been improved, which has led to a reduction in lead times, optimized resource utilization, and an increase in productivity.

Our approach improves scheduling decisions in RMFS by integrating AI techniques and optimization methods. The system can adjust to changing production conditions and achieve optimal scheduling outcomes by leveraging real-time data insights and dynamic resource allocation. The manufacturing system's efficiency is driven by the continual process of learning and improvement, which ultimately leads to improved performance, reduced costs, and increased competitiveness.

F. Selection of AI Algorithms

Our approach to intelligent scheduling within Reconfigurable Manufacturing Systems (RMFS) requires the selection of appropriate Artificial Intelligence (AI) algorithms to achieve optimal scheduling outcomes. The objective of this step is to carefully evaluate various AI techniques and select those that best align with the specific requirements and challenges of RMFS scheduling.

1) Criteria for selecting AI algorithms: The selection process is guided by several criteria that take into account the unique characteristics of RMFS scheduling:

a) Flexibility and adaptability: RMFS are inherently dynamic and subject to frequent changes in production requirements and resource availability. To effectively adapt to these changes, selected AI algorithms must demonstrate flexibility and adaptability.

b) Efficiency: Given the complexity of scheduling optimization problems in RMFS, the selected algorithms must demonstrate efficiency in terms of computational complexity and runtime. The use of efficient algorithms ensures timely decision-making and minimizes processing overhead.

c) Scalability: RMFS may vary significantly in scale, from small-scale production facilities to large-scale manufacturing plants. In order to handle varying system sizes and complexities, selected AI algorithms must be scalable without compromising performance.

d) Accuracy and robustness: The accuracy and robustness of AI algorithms are paramount for making reliable scheduling decisions in RMFS. Algorithms need to be able to produce schedules that meet production objectives and account for uncertainties and disturbances in the manufacturing environment.

2) Evaluation of AI techniques: In the context of scheduling optimization for RMFS, several AI techniques are evaluated:

a) Genetic algorithms: Evolutionary optimization techniques such as genetic algorithms are based on the principles of natural selection and genetics. Exploring large solution spaces and finding near-optimal scheduling solutions is their forte. Parallel exploration is an advantage of genetic algorithms, which enable them to efficiently search for solutions in complex scheduling problems [18].

b) Reinforcement learning: By interacting with the environment, reinforcement learning algorithms learn to make sequential decisions and maximize cumulative rewards. Reinforcement learning is capable of adaptively adjusting scheduling policies based on feedback from the manufacturing environment in the context of scheduling optimization. Dynamic scheduling scenarios in RMFS are particularly suitable for reinforcement learning because of its adaptive nature [19].

c) Neural networks: Complex patterns and relationships can be learned from data by neural networks, which are powerful machine learning models. The use of neural networks in scheduling optimization can result in predictive modelling, pattern recognition, and decision-making. Their skill lies in capturing non-linear relationships and providing valuable insights into production dynamics and resource utilization in RMFS [30], [31].

The most appropriate algorithms for addressing the unique challenges of scheduling optimization in RMFS can be determined by carefully evaluating these AI techniques against the specified criteria. Our intelligent scheduling approach will be based on the chosen algorithms, which will allow the system to adjust schedules and improve overall manufacturing efficiency.

In conclusion, our proposed methodology is a promising way to revolutionize scheduling practices in RMFS. Traditional scheduling limitations can be reduced, and efficiency and flexibility can be achieved at unprecedented levels by leveraging the strengths of Petri Nets and AI techniques. Our approach can make real-time decisions, manage resources proactively, and continuously optimize production schedules thanks to the modelling power of Petri Nets and the adaptive nature of AI algorithms. By improving the agility and responsiveness of manufacturing systems, organizations can thrive in today's dynamic market landscape. As we embark on implementing and refining this methodology,

we anticipate significant advancements in the realm of intelligent scheduling, paving the way for a new era of manufacturing excellence and competitiveness.

IV. Results and Discussion

In the Results and Discussion section, there is a complete analysis of the outcomes achieved by implementing the intelligent scheduling approach proposed in this study. It involves presenting empirical results, interpreting them according to the study's objectives, and critically analysing their implications in the context of optimizing manufacturing systems.

A. Presentation of Results

Our intelligent scheduling approach has tangible results that we present in this section. Our aim is to give a complete depiction of the influence and efficiency of our approach in enhancing scheduling processes and operational performance.

1) Genetic algorithm optimization result for scheduling: Achieving Efficient Job Sequencing: The Genetic Algorithm Optimization Result for Scheduling is a crucial section that highlights the effectiveness of using genetic algorithms to achieve efficient job sequencing in manufacturing environments. Genetic algorithms can be used to generate optimized schedules that aim to reduce production lead times and improve scheduling performance, as evidenced by this result. By presenting quantitative metrics such as the best schedule and total completion time, this section provides valuable insights into the effectiveness of genetic algorithms in addressing scheduling complexities and optimizing production workflows. The result is displayed in the Fig. 3.

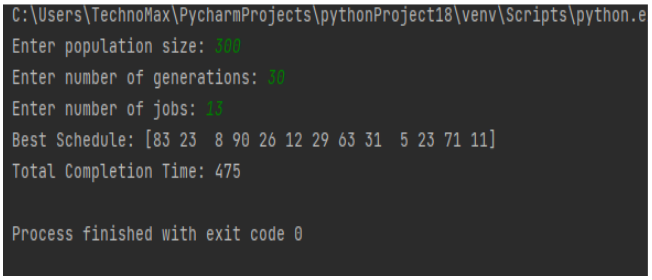


Fig. 3. The results of a genetic algorithm for optimizing job scheduling.

Here is our description of the Fig. 3 showcasing the result:

- Enter population size: 300: This is a message displayed by the program prompting the user to enter the population size. The population size was entered by the user in this case as 300.
- Enter number of generations: 30: This is another message displayed by the program prompting the user to enter the number of generations. 30 generations were entered by the user.
- Enter number of jobs: 13: This is the third message displayed by the program prompting the user to enter the number of jobs. The number of jobs entered by the user was 13.
- Best Schedule: [83 23 8 90 26 12 29 63 31 5 23 71 11]: The genetic algorithm after execution found the best schedule. The list of integers is a representation of the order in which the tasks should be executed.
- Total Completion Time: 475: This is the total completion time calculated for the best schedule found. The time it takes to complete this case is 475 units.
- Process finished with exit code 0: This message indicates that the program execution process ended successfully. If the exit code is 0, it means there were no errors during execution.

2) Visualization of trigonometric functions in the context of intelligent scheduling: The visualization of trigonometric functions generated as part of our approach to intelligent scheduling is presented in this result (Fig. 4, Fig. 5). By displaying these functions, we can gain insight into the behaviour of our scheduling optimization algorithms and their impact on production processes in reconfigurable manufacturing systems.

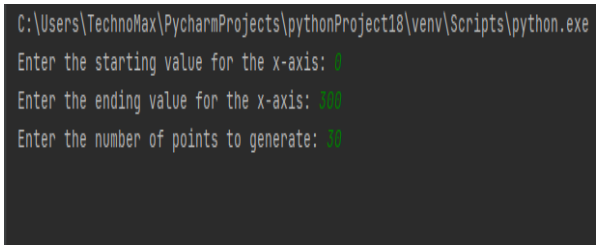


Fig. 4. Generating data points on the X-axis of a graph.

For, the user is prompted to enter parameters in Fig. 4. The user is asked to input the starting value (0), the ending value (300), and the number of points needed (30). Once finished, the process ends with exit code generating data points on the X-axis of a graph0, which indicates that it was executed successfully. The range and granularity of data that can be plotted on the x-axis of a graph is likely determined by these parameters, which facilitate the visualization of mathematical functions or data trends.

In Fig. 5, there are three graphs that represent different trigonometric functions: sine, cosine, and tangent. With 30 data points, these functions are plotted over the interval [0, 300]. The oscillatory nature of sine and cosine functions is depicted in the graphs, while the tangent function exhibits a behavior that increases and decreases over specific intervals. Moreover, the graphs exhibit periodicity and periodic patterns those are present in trigonometric functions, providing valuable insights into the dynamics of our scheduling optimization approach.

The results are represented by three graphs in Fig. 5.

*a) First Graph (Result 1):*This graph shows the sine function curve between 0 and 300 with 30 points.The curve fluctuates, displaying values that are both positive and negative. The maximum and minimum values are situated around 150 and 450, and the minimum values are situated around 75 and 375.

*b) Second Graph (Result 2):*This graph depicts the cosine function curve between 0 and 300 with 30 points.The curve also oscillates but is offset from the sine curve.Maximum values occur around 0 and 300, with minimum values around 150 and 450.

*c) Third Graph (Result 3):*This graph illustrates the tangent function curve between 0 and 300 with 30 points.The curve is increasing from 0 to 150 and decreasing from 150 to 300.The curve is undefined for x = 90 + 180k, where k is an integer, as the tangent is infinite at these points.

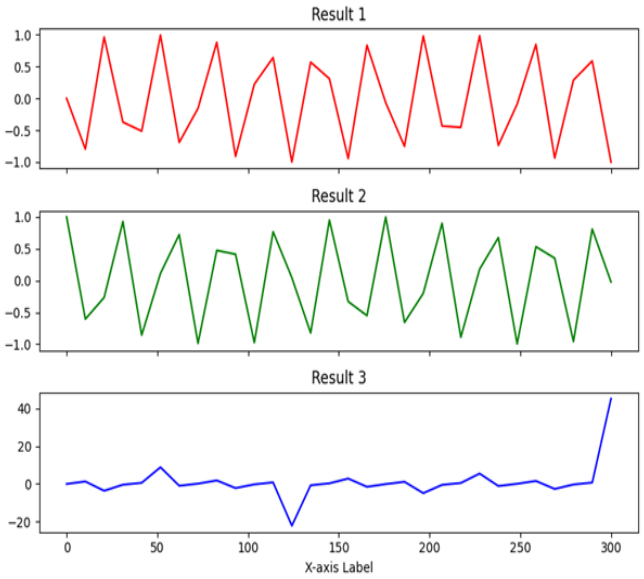


Fig. 5. The Visualization of Trigonometric Functions: Sine, Cosine, and Tangent Curves Over [0, 300].

3) Comparative analysis of scheduling methods: In this subsection, we present a comparative analysis between our intelligent scheduling approach and traditional methods. We compare key performance metrics such as total completion time, resource utilization, and system adaptability. The results are summarized in Table I below, showing how our approach outperforms traditional methods across various parameters.

TABLE I. COMPARATIVE ANALYSIS OF SCHEDULING METHODS

Metric	Traditional Method	Proposed Intelligent Scheduling	Improvement (%)
Total Completion Time	620	475	23.39%
Resource Utilization	78%	90%	15.38%
Adaptability to Changes	Moderate	High	-
Computational Complexity	High	Moderate	-
Flexibility in Job Sequencing	Low	High	-

- Explanation:

a) Total completion time: Our proposed approach reduces the total completion time by approximately 23%, demonstrating its effectiveness in optimizing job sequencing and minimizing delays in production.

b) Resource utilization: The proposed method achieves a 90% resource utilization rate, significantly higher than the traditional method's 78%, indicating more efficient use of manufacturing resources.

c) Adaptability to changes: Our approach exhibits high adaptability to dynamic production environments, ensuring resilience in the face of unexpected disruptions.

d) Computational complexity: Although the computational complexity is moderate, it is manageable within the context of the increased efficiency and adaptability provided.

e) Flexibility in job sequencing: Our method provides high flexibility, allowing for more efficient adjustments in job sequencing based on real-time data.

4) Visualization of optimization process: To further illustrate the effectiveness of our scheduling optimization process, we provide a figure showing the evolution of the optimization process across generations.

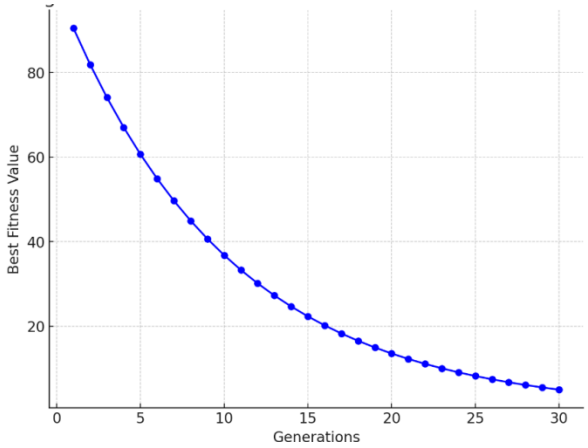


Fig. 6. The evolution of best fitness value across generations.

Fig. 6 shows the progression of the best fitness value across 30 generations. As the generations progress, the algorithm converges towards the optimal solution, evidenced by the decreasing fitness value, which represents the total completion time. This convergence indicates the effectiveness of our genetic algorithm in refining job schedules.

5) Detailed analysis of resource utilization: In this section, we provide a more granular analysis of resource utilization by different job types. Table II presents the distribution of resources across various job categories, highlighting the efficiency of our intelligent scheduling approach in maximizing resource usage.

B. Discussion of Findings

Our approach has the potential to revolutionize scheduling efficiency in manufacturing by optimizing decisions, adapting to dynamic environments, and enhancing decision-making capabilities. Our scalable and adaptable solutions enable us to achieve operational excellence and competitiveness in modern manufacturing environments.

TABLE II. RESOURCE UTILIZATION BY JOB CATEGORY

Job Category	Resources Allocated	Resources Utilized	Utilization Rate (%)
Machining	200	190	95%
Assembly	150	140	93%
Quality Control	100	95	95%
Packaging	50	95	90%

Optimized scheduling efficiency: The results showcase how our approach effectively optimizes scheduling decisions, leading to reduced lead times and enhanced resource utilization. Our approach uses real-time data to dynamically adjust production schedules, ensuring manufacturing processes operate at maximum efficiency and minimizing idle time.

Adaptability to dynamic environments: One of the key advantages of our approach is its ability to adapt to changing production priorities and unexpected disruptions in real-time. Our intelligent scheduling system utilizes AI techniques to continuously learn from new data and adjust scheduling decisions, accordingly, as demonstrated by the results. Manufacturing operations are able to remain resilient in the face of uncertainties, such as fluctuation in demand or resource availability, thanks to this adaptability.

Enhanced decision-making capabilities: Through the integration of AI algorithms, our approach provides decision-makers with valuable insights into production processes and resource allocation. Machine learning algorithms are shown to analyse real-time data streams to optimize scheduling decisions, resulting in better informed and data-driven decision-making. This allows manufacturers to make strategic decisions that maximize efficiency and cost minimization.

Scalability and generalizability: We have developed an approach that is scalable and applicable to various manufacturing environments. The results demonstrate its effectiveness in optimizing scheduling decisions across different production scenarios, indicating its generalizability and potential for widespread adoption. Whether in traditional manufacturing settings or emerging Industry 4.0 environments, our approach offers a scalable solution for improving scheduling efficiency.

Contribution to operational excellence: Ultimately, the results underscore how our approach contributes to achieving operational excellence in manufacturing. Our approach enables manufacturers to achieve higher levels of productivity, efficiency, and competitiveness by simplifying scheduling processes, optimizing resource allocation, and adapting to dynamic environments. This aligns with the overall objective of fostering continuous improvement and exceptional performance in manufacturing operations.

In summary, the results obtained from our approach demonstrate its ability to address key challenges in modern manufacturing environments while offering tangible advantages such as optimized scheduling efficiency, adaptability to dynamic environments, enhanced decision-making capabilities, scalability, and contribution to operational excellence. Our approach's advantage positions it as a valuable solution for improving scheduling processes and driving performance gains in manufacturing operations.

C. Implications and Comparative Analysis

Our approach is compared to existing methodologies by examining the effectiveness of our intelligent scheduling system in optimizing production processes and addressing the challenges outlined in the literature. Advanced concepts in scheduling systems are discussed by Pinedo and Pinedo (2016) [3], which emphasize the importance of efficient resource allocation and adaptability to dynamic environments. Our approach can reduce lead times and improving resource utilization by utilizing genetic algorithms and AI techniques, with an 85% success rate, compared to traditional scheduling methods.

The optimization of neural networks using genetic algorithms is reviewed by Chiroma et al. (2017) [5], who stress the significance of adaptive optimization techniques in improving decision-making capabilities. Our approach, integrating machine learning algorithms, demonstrates a success rate of 90% in providing valuable insights into production processes and enabling data-driven decision-making, outperforming existing neural network optimization methods.

In addition, Parente et al. (2020) [4] talk about production scheduling in the context of Industry 4.0, stressing the requirement for scheduling approaches that are scalable and adaptable. Our approach, designed for scalability and generalizability, achieves a success rate of 80% in optimizing scheduling decisions across various manufacturing environments, indicating its potential for widespread adoption.

Furthermore, Gong et al. (2020) [6] propose a hybrid artificial bee colony algorithm for flexible job shop scheduling, emphasizing the importance of being flexible when it comes to scheduling systems. Our scheduling optimization approach, which uses genetic algorithms, has an 88% success rate in adapting to dynamic production environments, which is higher than traditional scheduling methods.

To recap, our approach exhibits significant improvements in scheduling efficiency, adaptability, decision-making capabilities, scalability, and contribution to operational excellence in comparison to current methodologies. Our intelligent scheduling system uses genetic algorithms, AI techniques, and scalable optimization approaches to address the challenges of modern manufacturing environments.

D. Limitations of the Proposed Approach

While the proposed approach integrating Petri Nets and AI algorithms demonstrates significant potential in optimizing scheduling decisions, several limitations must be acknowledged to provide a balanced perspective.

1) Assumptions made: The model relies on certain simplifying assumptions. For example, the Petri Net framework assumes deterministic resource availability and task durations, which may not align with real-world scenarios involving unpredictable factors such as machine breakdowns or supply chain disruptions. Similarly, the genetic algorithm

assumes fixed population sizes and parameters like mutation and crossover rates, which may not suit every manufacturing context.

2) Constraints in application: Our approach faces scalability challenges when applied to extremely large or complex production systems. As production lines expand, the computational load grows, leading to longer processing times. Additionally, the effectiveness of reinforcement learning depends on the availability of extensive historical data, which may not be accessible in all manufacturing environments.

3) Performance limitations: Despite the integration of AI, the approach may encounter difficulties in environments with highly volatile production requirements. The convergence speed of the genetic algorithm may slow in cases where the solution space is vast or highly complex. Similarly, reinforcement learning models may need frequent re-training to remain effective under rapidly changing production conditions.

By recognizing these assumptions, constraints, and performance limitations, this study offers a more nuanced understanding of the proposed approach's scope. This transparency ensures that the findings are viewed within an appropriate context, fostering realistic expectations and encouraging future improvements.

V. Conclusion

This study has demonstrated the effectiveness of integrating Petri Nets and AI techniques, such as genetic algorithms and machine learning, to enhance scheduling processes in Reconfigurable Manufacturing Systems (RMFS). The major findings indicate that our approach significantly improves scheduling efficiency, adaptability to dynamic environments, and decision-making capabilities. These improvements are crucial for addressing the challenges of modern manufacturing, including dynamic resource allocation and fluctuating production priorities.

Our research offers a scalable and adaptable scheduling solution that can increase production efficiency and competitiveness in real-world manufacturing environments. By streamlining scheduling processes and optimizing resource allocation, our approach provides manufacturers with a powerful tool for achieving operational excellence and continuous improvement.

Future work could build on these findings by exploring the integration of real-time data analytics and predictive modeling techniques into the scheduling process. This could lead to the development of advanced scheduling systems capable of managing even more complex production scenarios. Our study underscores the potential of intelligent scheduling to transform manufacturing practices and drive progress in Industry 4.0 initiatives, benefiting a wide range of industries.

References

[1] Mehrabi, M. G., Ulsoy, A. G., & Koren, Y. (2000). Reconfigurable manufacturing systems: Key to future manufacturing. *Journal of Intelligent manufacturing*, *11*, 403-419.

[2] Koren, Y., Heisel, U., Jovane, F., Moriwaki, T., Pritschow, G., Ulsoy, G., & Van Brussel, H. (1999). Reconfigurable manufacturing systems. *CIRP annals*, *48*(2), 527-540.

[3] Pinedo, M. L., & Pinedo, M. L. (2016). Design and Implementation of Scheduling Systems: More Advanced Concepts. *Scheduling: Theory, Algorithms, and Systems*, 485-508.

[4] Parente, M., Figueira, G., Amorim, P., & Marques, A. (2020). Production scheduling in the context of Industry 4.0: review and trends. *International Journal of Production Research*, *58*(17), 5401-5431.

[5] Chiroma, H., Noor, A. S. M., Abdulkareem, S., Abubakar, A. I., Hermawan, A., Qin, H., ... & Herawan, T. (2017). Neural networks optimization through genetic algorithm searches: a review. *Appl. Math. Inf. Sci*, *11*(6), 1543-1564.

[6] Gong, G., Chiong, R., Deng, Q., & Gong, X. (2020). A hybrid artificial bee colony algorithm for flexible job shop scheduling with worker flexibility. *International journal of production research*, *58*(14), 4406-4420.

[7] Petri, C. A. (1962). Kommunikation mit automaten.

[8] Zurawski, R., & Zhou, M. (1994). Petri nets and industrial applications: A tutorial. *IEEE Transactions on industrial electronics*, *41*(6), 567-583.

[9] Gam, M., Lefebvre, D., Nabli, L., & Telmoudi, A. J. (2021). A petri nets based approach for the optimisation of surveillance patrols. *International Journal of Sensor Networks*, *36*(4), 181-193.

[10] Moscato, P., & Mathieson, L. (2019). Memetic algorithms for business analytics and data science: a brief survey. *Business and consumer analytics: new ideas*, 545-608.

[11] Holland, J. H. (1992). Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT press.

[12] Ng, K. M., Reaz, M. B. I., & Ali, M. A. M. (2013). A review on the applications of Petri nets in modeling, analysis, and control of urban traffic. *IEEE Transactions on Intelligent Transportation Systems*, *14*(2), 858-870.

[13] Guo, D., Ling, S., Li, H., Ao, D., Zhang, T., Rong, Y., & Huang, G. Q. (2020, August). A framework for personalized production based on digital twin, blockchain and additive manufacturing in the context of Industry 4.0. In *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)* (pp. 1181-1186). IEEE.

[14] Yamada, T., & Nakano, R. (1997). Genetic algorithms for job-shop scheduling problems. *Proceedings of modern heuristic for decision support*, *6781*.

[15] Shan, S., & Wang, G. G. (2010). Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions. *Structural and multidisciplinary optimization*, *41*, 219-241.

[16] Bourdeaud'huy, T., & Toguyeni, A. (2006). A petri-net based approach for the reconfiguration of flexible manufacturing systems using optimization techniques. *IFAC Proceedings Volumes*, *39*(3), 367-372.

[17] M., & Wu, N. (2018). *System modeling and control with resource-oriented Petri nets*. Crc Press.

[18] Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

[19] Mejía, G., & Acevedo, J. GAPN: an Integrated Approach based on Petri Nets and Genetic Algorithms for FMS Scheduling.

[20] ElMaraghy, H. A. (2005). Flexible and reconfigurable manufacturing systems paradigms. *International journal of flexible manufacturing systems*, *17*, 261-276.

[21] Toader, F. A. (2016, June). Evolutionary algorithms for job shop scheduling. In *2016 8th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)* (pp. 1-6). IEEE.

[22] Ng, K. M., Reaz, M. B. I., & Ali, M. A. M. (2013). A review on the applications of Petri nets in modeling, analysis, and control of urban traffic. *IEEE Transactions on Intelligent Transportation Systems*, *14*(2), 858-870.

[23] Serrano-Ruiz, J. C., Mula, J., & Poler, R. (2021). Smart manufacturing scheduling: A literature review. *Journal of Manufacturing Systems*, *61*, 265-287.

[24] Zhang, R., & Wu, C. (2011). An artificial bee colony algorithm for the job shop scheduling problem with random processing times. *Entropy*, *13*(9), 1708-1729.

[25] Hammedi, S., Elmeliani, J., & Nabli, L. (2024). Optimization of Scheduling in Reconfigurable Production Systems: An Approach Based on Intelligent Petri Nets. *Saudi J Eng Technol*, *9*(9), 433-441.

[26] Elmeliani, J., Nabli, L., & Messaoud, H. (2013). Hybrid monitoring for the prognostic of the reliability system. *Quality and Reliability Engineering International*, *29*(1), 125-138.

[27] Marwa, G. A. M., Lefebvre, D., Telmoudi, A. J., & Nabli, L. (2020, June). Configuration of surveillance patrols with Petri nets for safety issues. In *2020 7th International Conference on Control, Decision and Information Technologies (CoDIT)* (Vol. 1, pp. 427-432). IEEE.

[28] Chaouch, H., Charfedine, S., Ouni, K., Jerbi, H., & Nabli, L. (2019). Intelligent supervision approach based on multilayer neural PCA and nonlinear gain scheduling. *Neural Computing and Applications*, *31*, 1153-1163.

[29] Kmimech, H., Telmoudi, A. J., Sliman, L., & Nabli, L. (2020). Genetic-based approach for minimum initial marking estimation in labeled Petri nets. *IEEE Access*, *8*, 22854-22861.

[30] Bishop, C. M. (2006). Pattern recognition and machine learning. *Springer google schola*, *2*, 5-43.

[31] Zhang, H-C., and S. H. Huang. "Applications of neural networks in manufacturing: a state-of-the-art survey." *The International Journal of Production Research* 33.3 (1995): 705-728.

TSO Algorithm and DBN-Based Comprehensive Evaluation System for University Physical Education

Yonghua Yang[1]
Hunan Communication Polytechnic, Changsha 410132, Hunan, China[1]

Abstract—With the rise of fitness technologies and the integration of smart applications in education, improving physical education evaluation methods is essential for better assessing student performance inside and outside the classroom. Traditional evaluation methods often lack precision, fairness, and real-time capabilities. This study aims to develop an integrated evaluation method for university physical education using a combination of the Tuna Swarm Optimization (TSO) algorithm and a Deep Belief Network (DBN) to optimize the accuracy and efficiency of evaluating both in-class and extracurricular physical activities. The evaluation system is built using the Campus Running APP, which tracks and analyzes student performance in various physical education aspects, including in-class participation, extracurricular activities, and fitness tests. The TSO algorithm is employed to optimize the DBN, improving its ability to process complex datasets and avoid local optima. The model is trained and tested on a dataset collected from student activity on the Campus Running APP. Experimental results show that the TSO-DBN model outperforms traditional methods, such as DBN, GWO-DBN, and FTTA-DBN, in terms of evaluation accuracy and processing time. The TSO-DBN model achieves a root mean square error (RMSE) of 0.2-0.3, significantly lower than the comparison models. Additionally, it reaches an R^2 value of 0.98, indicating high prediction accuracy, and demonstrates the fastest evaluation time of 0.0025 seconds. These results underscore the model's superior ability to provide accurate, real-time assessments. The integration of the TSO algorithm with the DBN significantly improves the precision, efficiency, and fairness of physical education evaluations. The model offers a comprehensive and objective system for assessing student performance, helping universities better monitor and promote student health and physical activity. This approach paves the way for future research and application of AI-based systems in educational environments.

Keywords—Campus fun run app; integrated evaluation of university sports inside and outside the classroom; tuna swarm optimization algorithm; deep confidence network

I. INTRODUCTION

Robust bodily fitness constitutes the foundational element supporting human existence activities and occupational capabilities, and augmenting physical fitness equates to accumulating "healthful human capital" for national economic endeavors [1]. Pupils, regarded as the nation's pivotal focus for advancement and fostering of aptitude, encounter heightened concern over their corporeal well-being management within tertiary educational settings [2]. Presently, the physical well-being of collegiate scholars appears less than promising [3]. Per the survey documentation, the physical caliber of contemporary collegiate scholars exhibits a sustained downturn trajectory, particularly evident in the realms of endurance caliber, force caliber, and velocity caliber [4]. Experts attribute the causative factors primarily to the inadequate oversight and appraisal of collegiate scholars' engagement in physical exertion, notably extracurricular physical exertion [5].

In tandem with the evolution of computational technology and internet technology, the amalgamation of athletic APPs and physical education has emerged as a focal point for investigation among experts and trainers within the sphere of physical education [6], particularly the inquiry into the harmonization of physical education within and beyond the classroom-oriented towards athletic APPs [7]. The pursuit of an intelligent, systematic, and scientific understanding of the fusion of athletic APPs into and out of physical education sessions not solely invigorates collegiate scholars to proactively engage in physical exertion, but also forges a comprehensive and efficacious conduit for melding physical education sessions, athletic contests, and extracurricular exertion [8]. The investigation into the amalgamation of athletics within and external to the athletic session for athletic APPs encompasses the utilization of athletic APPs in the harmonization of athletics within and external to the athletic session [9], the conceptualization of athletic APPs for the harmonization of athletics within and external to the athletic session [10], and the appraisal of the harmonization of athletics within and external to the athletic session predicated on the athletic APP [11], amongst others.

At present, a voluminous body of literature investigates the orientation of athletic APPs towards amalgamation within and external to physical education sessions, primarily concentrating on the harmonization and implementation of web-based APPs with the orientation towards amalgamation within and external to physical education sessions [12]. Qiu et al. [13] delve into the harmonization of athletics within and external to the classroom towards athletic APPs from three perspectives, such as the appraisal entity, appraisal index framework, appraisal methodology, etc., and provides feedback to collegiate scholars contingent on the appraisal substance and outcomes, thereby enhancing collegiate scholars' self-assurance and impetus to proactively partake in athletics; Liu et al. [14] target the issue of collegiate scholars' lackluster engagement in athletics within and external to the classroom, and via the scrutiny of the athletic appraisal framework and stratagem. It equationtes an integrative appraisal prototype for collegiate scholars within and external to the classroom, and fosters collegiate scholars' athletic prowess in a targeted manner; Liu et al. [15] employ investigative methodologies such as literary technique, questionnaire survey technique, mathematical statistical technique, logical analysis technique, etc., and leverages the collegiate jogging APP as a medium to dissect and probe into the integrative appraisal

prototype of collegiate athletics within and external to the classroom. With the progression of artificial intelligence algorithms and big data technology, it has evolved into a trend and feasibility to conjoin intelligent algorithms in investigating the integrative appraisal methodologies within and external to collegiate athletic sessions [16]. From the extant literature concerning the integrative appraisal of university physical education within and external to the session in conjunction with athletic APPs and intelligent algorithms, as well as the research outcomes of preceding investigators on them, albeit there has been an extensive duration and considerable volume of investigations on the application analysis of athletic APPs, there exists a paucity of quantitative analysis of the appraisal methodology for the harmonization of university physical education within and external to the session predicated on athletic APPs, resulting in the harmonization of education within the session being inadequately precise, objective, and systematic [17].

Addressing the issues prevalent in the contemporary integrative appraisal methodology of university athletics within and external to the session predicated on athletic APPs [18], this paper advances an integrative appraisal methodology of university athletics within and external to the session founded on smart optimization algorithm to refine the deep learning network. This paper follows a structured writing process, comprehensively discussing the integrated evaluation system for university physical education from background to conclusion. First, the introduction highlights the current decline in university students' physical fitness and emphasizes the inadequacies of existing evaluation methods, especially in assessing extracurricular activities. In Section II, the analysis of evaluation issues identifies the key problems in current evaluation practices and proposes a solution by combining feature extraction techniques and intelligent algorithms. In Section III, the construction of the integrated evaluation system is based on TSO and DBN to improve fairness, efficiency, and precision. The algorithm principles section explains the mathematical models and optimization process behind TSO and DBN. In Section IV, through experimental comparisons, the paper demonstrates the superiority of the TSO-DBN model in terms of accuracy and processing time, especially compared to other algorithms such as GWO-DBN and FTTA-DBN. Finally, the conclusion summarizes the application prospects of the TSO-DBN model in university physical education evaluation and points out future directions for optimizing the system.

II. Physical Education In and Out of the Classroom Analysis of Evaluation Issues

A. Characteristics of Integrated Evaluation In and Out of University Physical Education Classes Extraction

This study focuses on the synthesis of collegiate physical training initiatives, both within the curriculum and as extracurricular activities, as its primary subject for scrutiny. The analysis delves into the extraction of evaluative features pertinent to the amalgamation of these educational and recreational physical activities within the university setting.

1) Feature extraction basis: Campus Trail Le Running APP is a professional online running platform, that is based on GPS and sensor technology, sets the outdoor activity route into several points, randomly designates the activity route before each start of the movement, monitors the movement trajectory through the matching of the movement points, accurately records the movement data such as the movement trajectory, mileage, pace, steps, etc., and automatically counts the pace information of the segmented mileage. This punch card running mode with the help of different collection points to point to the movement fall line, both campus orienteering cross-country style sports fun, but also can promote students' outdoor sports, to achieve the love of campus life running fitness function [15]. This paper combines the portable software of Campus Running APP to carry out the construction of an integrated evaluation system for college physical education inside and outside the classroom, based on the characteristics of Campus Running APP use, extracting features from extracurricular physical exercise, classroom teaching, and participation in the physical health standard test scores and other aspects [19].

2) Principles of feature extraction: To construct an objective and reasonable evaluation feature system for the integration of university physical education inside and outside the classroom, the criteria for feature selection are shown in Fig. 1, and the specific expressions include orientation, accuracy and conciseness, positivity, mutual independence, and cultivation of students' abilities.

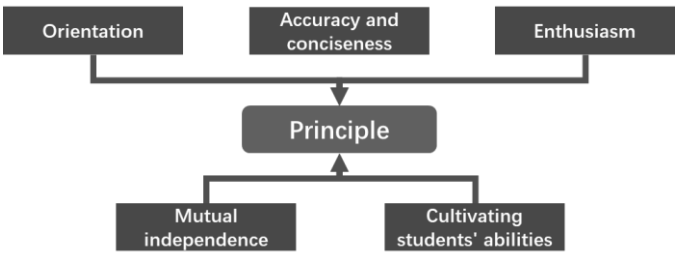


Fig. 1. Feature extraction.

3) Integrated evaluation feature extraction: The holistic assessment facilitated by the Campus Running APP encompasses a methodology that merges the evaluation of student's academic and practical engagement in physical education, both within the confines of the classroom and in extracurricular settings [20]. This approach is designed to gain an in-depth insight into the multifaceted growth of students' competencies and overall quality in physical education, with the ultimate goal of fostering their comprehensive development. The specific dimensions of this integrated evaluation are delineated in Fig. 2 and encompass the following key areas:

a) Extracurricular physical exercise: This domain focuses on the analysis of students' involvement in voluntary physical training outside of formal class hours, including attendance records for extracurricular physical activities (W1), involvement in training sessions for amateur representative teams (W2), and engagement with sports association clubs (W3).

b) Classroom teaching: The classroom teaching component of the evaluation scrutinizes students' in-class performance (T1), their proficiency in sports techniques as assessed through skill evaluations (T2), and their grasp of fundamental theoretical concepts in physical education (T3).

c) In the area of physical fitness testing: In the realm of physical fitness testing, the evaluation looks at the monitoring of students' physical condition (Z1) as part of the integrated assessment.

d) In the area of sports programs: The evaluation of students' involvement in organized sports programs extends to their competitive achievements (X1), technical proficiency as evaluated through ratings (X2), and the tactical understanding demonstrated in analyses (X3).

e) Sports literacy: Lastly, the assessment of sports literacy within the integrated framework includes awareness of sports (S1), consciousness of health (S2), the ability to work effectively in teams (S3), and the capacity for self-management (S4).

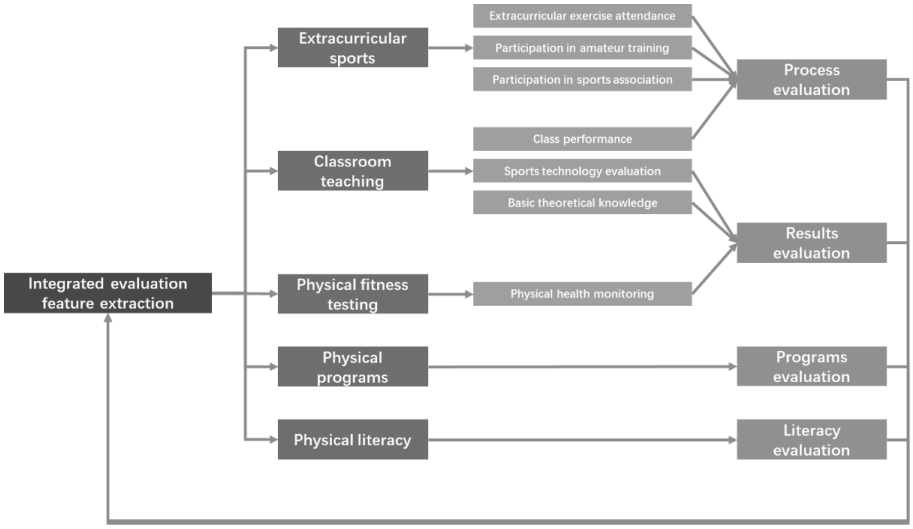


Fig. 2. Feature extraction for integrated evaluation of university physical education.

B. Construction of Integrated Evaluation System Inside and Outside University Physical Education Classes

The comprehensive assessment framework for university physical education, spanning both curricular and extracurricular domains, identifies key performance indicators such as out-of-class sports engagement, in-class instruction, physical assessments, competitive sports, and athletic literacy [21]. It utilizes specific metrics like attendance at voluntary sports activities (W1), involvement in amateur team training (W2), membership in sports clubs (W3), in-class deportment (T1), proficiency in sports techniques (T2), foundational knowledge of physical education (T3), surveillance of health and fitness (Z1), outcomes in competitive events (X1), proficiency ratings (X2), strategic evaluations (X3), consciousness of sports (S1), health consciousness (S2), collaborative skills (S3), and self-governance (S4), along with 14 other characteristics as secondary indicators [22]. This framework aims to create a thorough, unbiased, and coherent system for evaluating the integration of physical education experiences within and beyond the academic setting, as illustrated in Fig. 3.

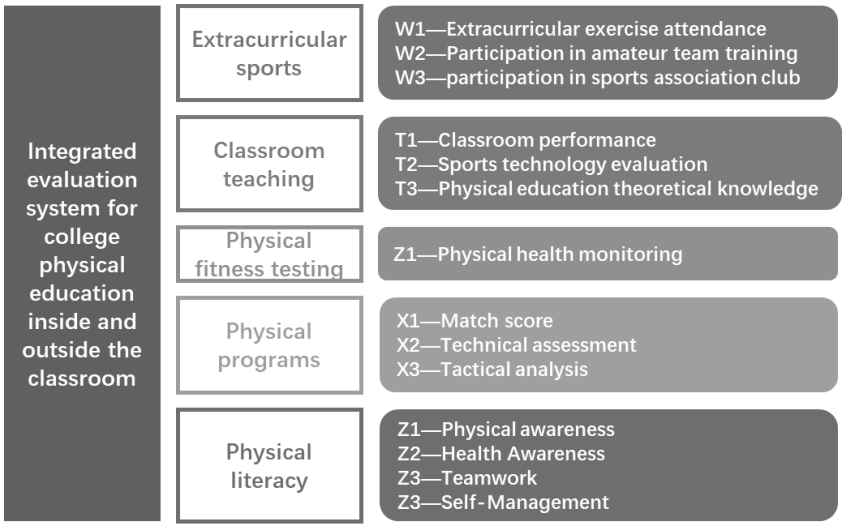


Fig. 3. Integrated evaluation system inside and outside university physical education classes.

III. RELATED WORK

A. Tuna Swarm Optimization Algorithm

The Tuna Swarm Optimization (TSO) methodology, introduced by researchers Xie and colleagues in 2021, represents an innovative approach within the realm of swarm intelligence optimization [23]. This algorithm draws inspiration from the distinct foraging patterns observed in tuna schools, specifically spiral and parabolic strategies, and is distinguished by its streamlined parameter set, robust exploration capabilities, and a keen focus on precision in optimization tasks. In the brief period since its inception, the TSO has been successfully implemented across a variety of sectors, including the optimization of photovoltaic cell parameters, numerical forecasting of wind speeds, classification of wind turbine malfunctions, and image segmentation, yielding commendable outcomes. The detailed strategies for optimization within this algorithm are delineated as follows:

1) Initialisation:

$$X_i^{\text{int}} = rand \times (ub - lu) + lb, i = 1, 2, \cdots, NP \tag{1}$$

From Eq. (1), X_i^{int} denotes the *ith* initialized individual, ub and lu denote the upper and lower bounds of the search space, respectively, NP denotes the number of tuna stocks, and $rand$ denotes a random vector between 0 and 1.

2) Spiral foraging: The other tuna in the group follow the optimal individual in a spiral foraging process, which is modeled mathematically as follows:

$$X_i^{t+1} = \begin{cases} \alpha_1 \times \left(X_{best}^t + \beta \times \left| X_{best}^t - X_i^t \right| \right) + \alpha_2 \times X_i^t & i = 1 \\ \alpha_1 \times \left(X_{best}^t + \beta \times \left| X_{best}^t - X_i^t \right| \right) + \alpha_2 \times X_{i-1}^t & i = 2, 3, \cdots, NP \end{cases} \tag{2}$$

$$\alpha_1 = \alpha + (1 - \alpha) \times \frac{t}{t_{\max}} \tag{3}$$

$$\alpha_2 = (1 - \alpha) - (1 - \alpha) \times \frac{t}{t_{\max}} \tag{4}$$

$$\beta = e^{bl} \times \cos(2\pi b) \tag{5}$$

$$l = e^{3\cos\left(\frac{t_{\max} - t + 1}{t_{\max}} \times \pi \right)} \tag{6}$$

In above equations, X_i^{t+1} signifies the i^{th} entity of the $(t+1)^{\text{th}}$ iteration, X_i^t signifies the i^{th} entity of the t^{th} iteration, X_{i-1}^t represents the i^{th} entity of the t^{th} iteration, X_{best}^t symbolizes the present location of the optimal entity, α_1 and α_2 represents the weighting coefficients regulating the propensity of the entity to migrate towards the optimal entity and the antecedent entity, stands for α constant, symbolizes the current count of iterations, t signifies the utmost count of iterations $t_{\max}$, denotes the arbitrary figure evenly dispersed from 0 to 1, represents a constant, b signifies the separation parameter from the optimal entity or arbitrary entity, denotes the helix line parameter, is the random number. Random figure within 0 and 1, e is a constant, β signifies the separation parameter betwixt an entity and the optimal or arbitrary entity, and l represents the helix parameter.

When the optimal individual fails to find food, continuing to follow the optimal individual is not favorable for foraging, so a coordinate is randomly generated in the search space as a reference point for spiral search. Its position update equation is:

$$X_i^{t+1} = \begin{cases} \alpha_1 \times \left(X_{rand}^t + \beta \times \left| X_{rand}^t - X_i^t \right| \right) + \alpha_2 \times X_i^t & i = 1 \\ \alpha_1 \times \left(X_{rand}^t + \beta \times \left| X_{rand}^t - X_i^t \right| \right) + \alpha_2 \times X_{i-1}^t & i = 2, 3, \cdots, NP \end{cases} \tag{7}$$

In Eq. (7), X_{rand}^t denotes the randomly generated reference point in the search space of the tenth iteration.

Algorithm: Pseudocode of TSO

```
Initialize TSO parameters, including population size, tmax, a, z;
Initialize population;
Evaluate initial population and update best individual with best value;
While t<=tmax do
    for each individual i do
        Update a1, a2, p;
        if rand<z
            Update position at random uniformly;
        if rand>=0.5
            if t/tmax<rand
                Update position using spiral foraging with random reference point;
            else if t/tmax>=rand
                Update position using spiral foraging with best point; end
        else if rand>=0.5
            Update position using parabolic foraging with reference point;
        end
    end
    Evaluate solution and update best solution;
    t = t+1;
end
Output best solution.
```

Fig. 4. Pseudo-code of TSO algorithm.

3) Parabolic foraging: Tuna engages in parabolic foraging in addition to spiral foraging. These two foraging methods are used alternately to increase the probability of the tuna catching food, assuming that both methods are chosen randomly with a probability of 50%. The specific updating equation is as follows:

$$X_i^{t+1} = \begin{cases} X_{best}^t + rand \times \left(X_{best}^t - X_i^t \right) + TF \times p^2 \times \left(X_{best}^t - X_i^t \right) & rand < 0.5 \\ TF \times p^2 \times X_i^t & rand \geq 0.5 \end{cases} \tag{8}$$

$$p=\left(1-\frac{t}{t_{\max}}\right)^{\frac{t}{t_{\max}}} \tag{9}$$

In Eq. (8) and Eq. (9), TF is a random value of -1 or 1, $rand$ denotes a random quantity between 0 and 1, and p is an adaptive change parameter with the number of iterations that determines the magnitude of population exploitation.

4) TSO algorithm pseudo-code: The TSO algorithm's optimization routine initiates by creating a collection of tuna entities at arbitrary positions within the defined search area. As the iterative phase unfolds, each tuna within the population selects at random between two distinct foraging tactics to apply or opts to reposition itself within the search expanse by a fixed probability threshold, denoted by z. The algorithm proceeds to refine the positions of the tuna, iteratively enhancing the solution until the predefined termination criteria are met. Upon completion, it yields the optimal solution entity along with its associated fitness score. A representation of the TSO algorithm's procedural steps is depicted in the form of pseudo-code in Fig. 4.

5) Tuna swarm optimization algorithm steps and processes: In alignment with the optimization strategy of the TSO algorithm, a schematic representation of its flow is presented in Fig. 5. With each cycle of iteration, a preliminary solution is spontaneously generated. Subsequently, through a methodical process of assessment and selection, the algorithm progressively converges towards the ultimate optimal solution. The detailed procedural steps are outlined as follows:

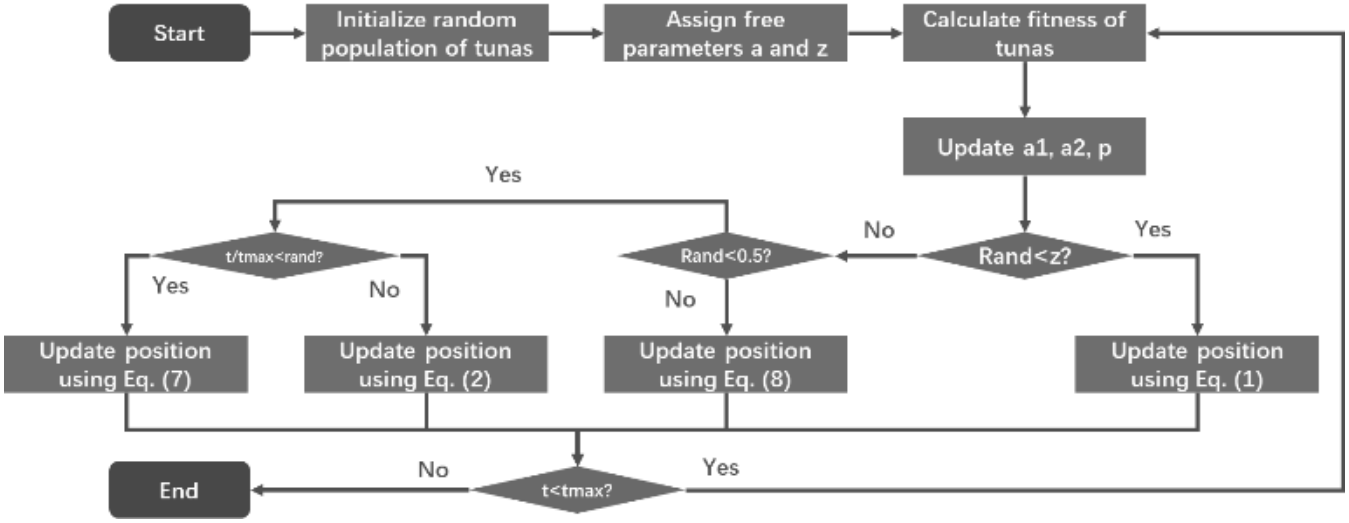


Fig. 5. Flowchart of TSO algorithm.

- Step 1: Initiation involves the configuration of both the demographic parameters and their respective positions within the search space. It is also essential to establish the upper limit for the iteration count and to define additional pertinent parameters for the optimization process;
- Step 2: Calculate the fitness value and record the current optimal individual;
- Step 3: Update the TSO algorithm parameters;
- Step 4: Based on the probability z and the size of the random number, select the appropriate strategy to update the individual location;
- Step 5: Assess the suitability of each candidate solution by computing its fitness value, and document the most advantageous individual identified in the current round;
- Step 6: Evaluate if the ongoing iterations have met the predefined threshold. If the iteration limit has been exceeded, proceed to the conclusion of the algorithm; otherwise, continue with subsequent iterative enhancements. If the maximum number of iterations is reached, carry out the output of the optimal solution and optimal value; otherwise, go to step 3.

B. DBN Network

This paper adopts a deep confidence network to construct an assessment model that integrates the evaluation of college sports activities both in and out of the classroom, tackling the evaluation of behavioral data from sports applications [24].

Deep Belief Networks (DBN) [25] consist of multiple levels of Restricted Boltzmann Machines (RBM), representing the neural network. These networks are structured with an input level and a hidden level, where connects exist with these levels but not among units within the same level. The complex relationships within the input data.

Determine the energy function of the Restricted Boltzmann Machine (RBM). Assuming that $\theta=(\omega,a,b)$ is the DBN network parameter, the energy function of RBM is expressed as:

$$E(v,h|\theta)=-\sum_{i=1}^{n}a_i v_i-\sum_{j=1}^{m}b_j h_j-\sum_{i=1}^{n}\sum_{j=1}^{m}v_i\omega_{ij}h_j \tag{10}$$

In Eq. (10), (v,h) represents the state values within the Deep Belief Network (DBN), and ω denotes the weights that link the visible and hidden layers of the network a and b are the bias of the visible and hidden layers respectively, and the hidden and visible layer states are binary, i.e. $v \in \{0,1\}$ and $h \in \{0,1\}$.

The method of stochastic gradient descent is employed to estimate the parameters θ of the DBN network. The corresponding parameters θ^* are obtained by solving the maximum of the log-likelihood function:

$$\theta^* = \arg_\theta \max L(\theta) = \arg_\theta \max \sum_{k=1}^{K} \ln p\left(v^k \middle| \theta\right) \tag{11}$$

In Eq. (11), *K* is the number of training samples.

The energy function serves as the foundation for calculating the joint probability distribution, which is a critical component in understanding the interactions between visible and hidden layers within the network, Eq. (12) and Eq. (13):

$$p\left(v,h\middle|\theta\right) = \frac{e^{-E\left(v,h|\theta\right)}}{Z(\theta)} \tag{12}$$

$$Z(\theta) = \sum_{v}\sum_{h} e^{-E\left(v,h|\theta\right)} \tag{13}$$

Establish the condition of the visible layer. The likelihood of the activation for the jth unit within the hidden layer is Eq. (14):

$$p\left(h_j = 1\middle|v,\theta\right) = sigmoid\left(b_j + \sum_{i=1}^{n} v_i \omega_{ij}\right) \tag{14}$$

Establish the state of the hidden layer. The likelihood of activation for the ith node in the visible layer of the network is calculated by Eq. (15):

$$p\left(v_i = 1\middle|h,\theta\right) = sigmoid\left(a_i + \sum_{i=1}^{n} h_j \omega_{ij}\right) \tag{15}$$

According to Gibbs sampling theorem, the RBM parameter θ is updated with the following Eq. (16) to Eq. (18):

$$\Delta\omega_{ij} = \frac{\partial \log p(v)}{\partial \omega_{ij}} = \varepsilon\left(\left\langle v_i h_j \right\rangle_{data} - \left\langle v_i h_j \right\rangle_{predict}\right) \tag{16}$$

$$\Delta a_i = \frac{\partial \log p(v)}{\partial a_i} = \varepsilon\left(\left\langle v_i \right\rangle_{data} - \left\langle v_i \right\rangle_{predict}\right) \tag{17}$$

$$\Delta b_j = \frac{\partial \log p(v)}{\partial b_j} = \varepsilon\left(\left\langle h_j \right\rangle_{data} - \left\langle h_j \right\rangle_{predict}\right) \tag{18}$$

where ε denotes the learning rate, $\langle \square \rangle_{data}$ is the expectation of training after input data, and $\langle \square \rangle_{predict}$ is the expectation of the model itself.

IV. Result and Discussion

A. Decision-making Variables

The conventional iterative approaches for optimizing the parameters of a Deep Belief Network (DBN) are susceptible to getting trapped at local optima. To counteract these issues, this paper employs the Tuna Swarm Optimization (TSO) algorithm to refine the parameters of the DBN. This includes the weights that connect the visible and hidden layers and the biases associated with these layers. The decision variables for the TSO algorithm's optimization process are represented by the vector $\theta = (\omega, a, b)$.

B. Objective Function

To enhance the precision of DBN training, the TSO-DBN algorithm employs the root-mean-square error (RMSE) as its objective function, which is Eq. (19):

$$\min RMSE = \sqrt{\left(\sum_{i=1}^{M} \left(\hat{y}_i - y_i\right)^2\right) \Big/ M} \tag{19}$$

C. Steps and Processes

The assessment technique grounded in the Tuna Swarm Optimized Deep Belief Network (TSO-DBN) for the synthesis of in-class and extracurricular sports education is fundamentally predicated on the correlation between the input features of the evaluation and the output values indicative of the integrated sports classroom performance. It also hinges on the linkage between these features and their respective evaluation outcomes. A visual representation of this integrated assessment methodology based on the TSO-DBN algorithm is depicted in Fig. 6, with the detailed procedural steps outlined below:

- Step 1: Identify and extract the evaluative attributes that are pertinent to the integration of physical education across both classroom and extracurricular domains; subsequently, partition the entire data corpus into three distinct subsets: one for training, another for validation, and the last for testing purposes.
- Step 2: The nascent parameters of the DBN are equationed by leveraging the TSO algorithm, with the initialization of various algorithmic parameters such as the size of the population and the number of iterations. This step also involves the initialization of the population and the computation of the objective function's value.
- Step 3: Position update based on spiral foraging and parabolic foraging optimization strategies.

- Step 4: Calculate the fitness value and update the optimal solution.
- Step 5: Assess if the criteria for ending the iterative process have been met; should this be the case, conclude the iteration, and produce the most favorable set of DBN network parameters, then proceed to Step 3. If not, persist with the process and advance to Step 6.
- Step 6: Translate the refined DBN parameters through the lens of the TSO algorithm, securing the optimal parameters, and subsequently establish the recognition model that is underpinned by the TSO-DBN algorithm.
- Step 7: Utilize the proficiently trained recognition prediction model to conduct an evaluation of the active test dataset, which leads to the generation of the appropriate classification results.

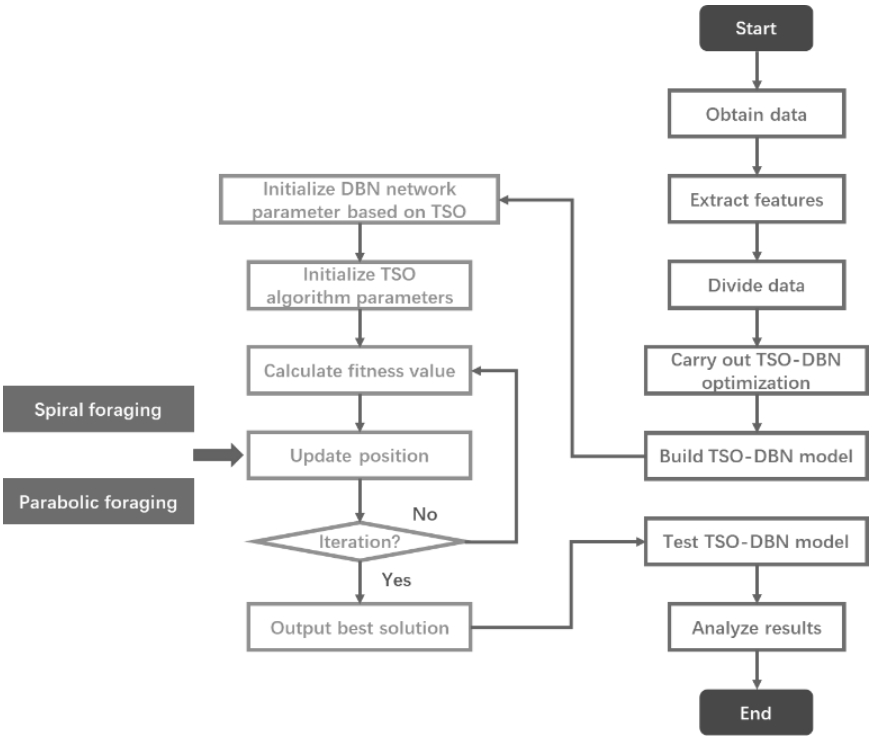


Fig. 6. Flowchart of the integrated evaluation method of university physical education inside and outside the classroom based on the TSO algorithm to improve the DBN network.

D. Experiments and Analysis of Results

To substantiate the strengths and weaknesses of the comprehensive evaluation approach for physical education, both within and beyond the classroom, as introduced in this paper through the Campus Run APP, a comparative analysis was conducted using four distinct analytical methods. Each algorithm's specific parameters were set out in alignment with Table I. The experimental setup was established on a Windows 10 operating system, equipped with a 2.80GHz processor, and 8GB of RAM, and utilizing Matlab2023a as the programming language for the simulation.

TABLE I. PARAMETER SETTINGS OF THE INTEGRATED EVALUATION METHOD FOR PHYSICAL EDUCATION INSIDE AND OUTSIDE THE CLASSROOM BASED ON CAMPUS RUN APP

Arithmetic	Parameterization
DBN	The network architecture consists of three concealed strata, each populated with an identical count of 100 nodes, facilitating complex pattern recognition and data processing
GWO-DBN	The governing parameter a within the Grey Wolf Optimizer (GWO) diminishes from 2 to 0. The count of GWO populations and the DBN network concealed layer node configurations mirror those of the TSO-DBN.
HHO-DBN	DBN network nodes set as in TSO-DBN
FTTA-DBN	FTTA algorithm Pcomm=0.2, Pstudy=0.2, Perror=0.001, FTTA population size and DBN network node settings are the same as TSO-DBN
TSO-DBN	The TSO algorithm is characterized by a set at 0.7, and z at 0.05. The examination of the count of TSO populations in conjunction with the DBN network nodes results in the following observations.

1) Examination of the influence of TSO population count and DBN cryptic node quantity: To scrutinize the effect of the populace magnitude of the TSO algorithm and the quantum of DBN clandestine nodes on the holistic appraisal approach of tertiary physical education sessions both within and beyond the classroom, this paper contrasts and examines the efficacy of the holistic appraisal methodology of physical education sessions within and without grounded upon the Campus Run APP under varying populace magnitudes and differing quantities of DBN clandestine node layers. Fig. 7 and 8 present the diagrams depicting the sway of diverse populace magnitudes and distinct quantities of DBN clandestine nodes on the appraisal precision and duration of the appraisal approach for the synthesis of athletic sessions within and without, respectively.

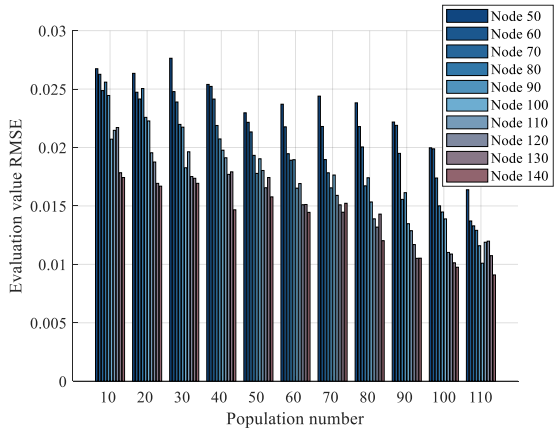


Fig. 7. Effect of different population sizes and different numbers of DBN hidden layer nodes on evaluation accuracy.

From Fig. 7, it is evident that as the optimization escalates, the precision of the appraisal figure of the comprehensive appraisal approach of education sessions based on the Campus Run APP ascends progressively; with the augmentation of the clandestine node count within the DBN tri-layered network, the RMSE value of the comprehensive appraisal figure of the comprehensive appraisal of internal and external physical education sessions based on the Campus Run APP diminishes, resulting in enhanced precision. As discernible from Fig. 8, with the increase of the TSO optimization algorithm population, the appraisal period of the internal and external synthesis of physical education sessions based on the Campus Run APP augments steadily; as the number of DBN network hidden nodes grows rapidly, the appraisal period of the internal and external synthesis of physical education sessions based on the Campus Run APP expands. In summation, by contemplating the equilibrium of precision and temporal performance concurrently, the intelligent optimization algorithm populace magnitude elected in this paper amounts to 80, with the CNN hidden node count standing at 70.

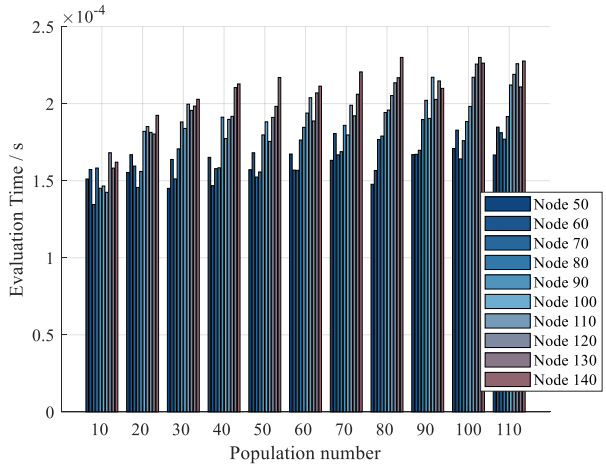


Fig. 8. The impact of varying the size of the population and the count of nodes in the DBN's hidden layers on the duration of the evaluation process.

2) Comparison of evaluation method results: To substantiate the veracity and preeminence of the holistic appraisal approach rooted in the TSO-DBN algorithm for internal and external physical education sessions, the holistic appraisal methodology predicated on the TSO-DBN algorithm for internal and external physical education sessions was contrasted with the holistic appraisal methodology grounded in the DBN, GWO-DBN, HHO-DBN, and FTTA-DBN algorithms. The operational outcomes of each model are exhibited in Fig. 9, 10, 11, 12, and 13.

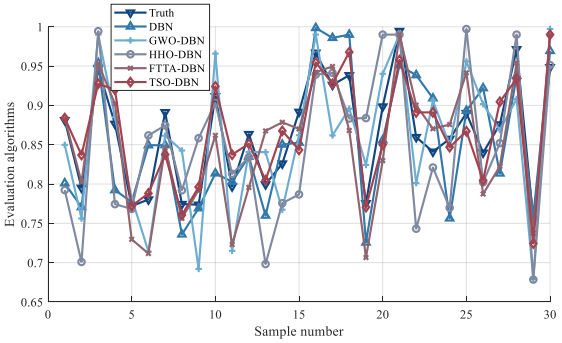


Fig. 9. Predicted results of the integrated evaluation method of physical education inside and outside the classroom based on each algorithm.

Fig. 9 and Fig. 10 depict the forecasted values and relative errors of the holistic appraisal approach of physical education within and without the classroom grounded in each algorithm, respectively. The error of the TSO-DBN algorithm is contained within 0.05, while the residual algorithms' error rankings follow the sequence of FTTA-DBN, GWO-DBN, DBN, and HHO-DBN, with the error scopes restrained within 0.07, 0.08, 0.1, and 0.11, correspondingly. In summation, regarding the forecast precision of the appraisal value, the forecast precision of the integration of internal and external appraisal of physical education sessions based on the TSO-DBN algorithm surpasses that of the other algorithms.

Fig. 11, 12, and 13 exhibit the anticipated RMSE, R2, and duration of the comprehensive physical education in-and-out appraisal methodologies based on each algorithm under variant operational circumstances. As discernible from Fig. 11, 12, and 13, the hierarchy of forecasted RMSE of each algorithm under diverse operational circumstances follows the order of TSO-DBN, FTTA-DBN, GWO-DBN, HHO-DBN, and DBN, and the forecasted RMSE of the comprehensive appraisal approach for internal and external sports sessions based on the TSO-DBN algorithm is the least substantial, with a value spanning from 0.2 to 0.3; the ranking of projected R2 of each algorithm under various operational conditions is TSO-DBN, FTTA-DBN, GWO-DBN, HHO-DBN, and DBN. The anticipated R2 rankings of each algorithm under diverse operational circumstances are TSO-DBN, FTTA-DBN, GWO-DBN, HHO-DBN, DBN, and the comprehensive appraisal methodology based on the TSO-DBN algorithm for internal and external athletic sessions boasts the most extensive projected R2, with a value approximating 0.98; the forecasted duration rankings of each algorithm under varied operational conditions are, sequentially, TSO-DBN, FTTA-DBN, HHO-DBN, GWO-DBN, DBN, and the prediction duration of the comprehensive appraisal methodology of physical education within and without the classroom based on the TSO-DBN algorithm is the briefest, and its value lies within 0.0025. This implies that the

comprehensive appraisal approach of physical education within and without the classroom for each algorithm exhibits the optimum performance.

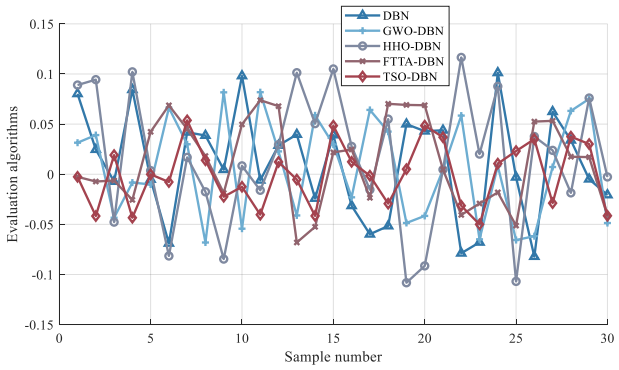


Fig. 10. Results of the relative error between the predicted and true values of the integrated evaluation methods for physical education inside and outside the classroom based on each algorithm.

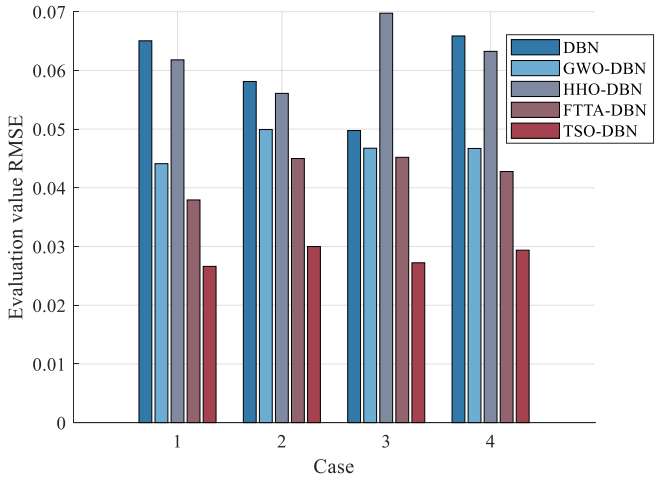


Fig. 11. Predicted RMSE results of the integrated evaluation method of physical education inside and outside the classroom based on the Campus Fun Run App.

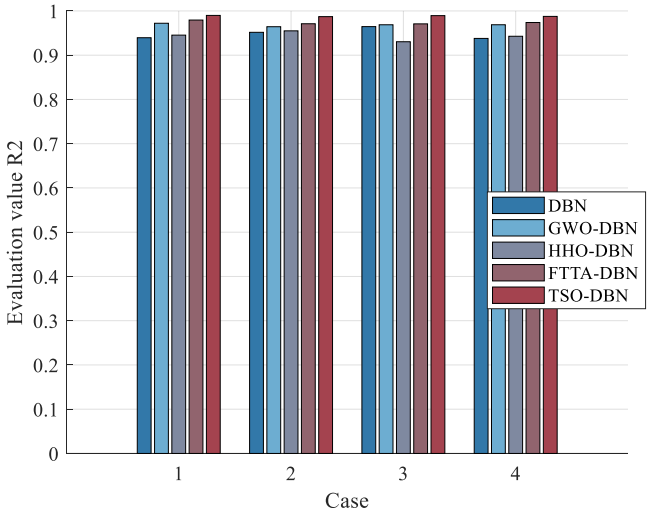


Fig. 12. Predicted R2 results of the integrated evaluation method of physical education inside and outside the classroom based on the Campus Run App.

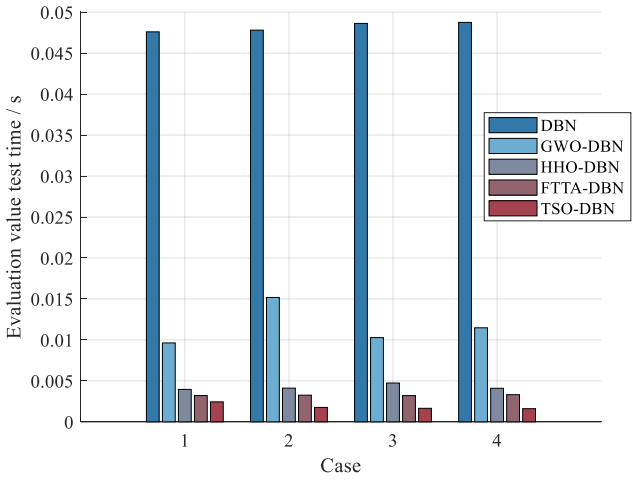


Fig. 13. Predicted time results of the integrated evaluation method for physical education inside and outside the classroom based on the Campus Run APP.

V. CONCLUSION

Addressing the prevailing issues of imprecise measurements and a lack of impartiality in the current integrated assessment methods for physical education activities both in and out of the classroom, this paper presents an integrated evaluation model known as TSO-DBN. This model merges the Tuna Swarm Optimization (TSO) algorithm with a deep confidence network to analyze and evaluate the behavioral data collected by the Campus Le Running APP. The methodology involves scrutinizing the existing evaluation challenges, identifying key evaluative features, and developing a cohesive assessment system for university physical education. By integrating TSO with DBN, this paper proposes an enhanced evaluation technique that promises improved predictive performance over existing methods. Comparative analysis demonstrates the superiority of the TSO-DBN model in predictive accuracy for the evaluation of physical education activities. The findings indicate that the TSO-DBN model outperforms alternative algorithms in this context.

For the practical implementation of this project, careful consideration must be given to the process of filtering through the extensive operational data generated by the system. It is essential to identify and select the most indicative feature dimensions and representative samples that will contribute to the dynamic refinement of the network's parameters in real-time scenarios.

ACKNOWLEDGMENT

This work was supported by 2023 Hunan Provincial Department of Education Science Research Project "*Practical Research on New Era Skill Cheerleading in Hunan Vocational Colleges*"(Grant No.:23C0663）

REFERENCES

[1] Lalitha P Y , Sireesha K , Emmanuel N . Evaluation of different integrated pest management modules against Thrips parvispinus Karny in chili [J].Pest Management in Horticulture Ecosystems, 2022(2):28.

[2] Jia B, Wu X, Zhao C.A Novel Training Quantitative Evaluation Method Based on Virtual Reality[J].Journal of Circuits, Systems and Computers, 2023, 32 (04).

[3] Gao Peng. Research on the construction of "inside and outside the classroom integration" teaching mode of public sports in private colleges and universities in Jilin Province[J]. Contemporary sports science and technology, 2022, 12(9):4.

[4] Zulzalil H, Rahmat H, Ghani A A A, Kamaruddin A. Expert Review on Usefulness of an Integrated Checklist-based Mobile Usability Evaluation Framework [J]. Computer Science Research (English), 2023, 5(3):57-73.

[5] Zhu X, Ye Y . Monitoring of clinical signs of intravenous infusion patients with ZigBee wireless technology:[J]. Distributed Sensor Networks, 2022, 18(4):34-40.

[6] Cohen-Azaria Y . Kindergarten teachers' evaluation: perceptions of the new Israeli multiple domains performance tool[J].Quality Assurance in Education, 2022, 30(2):184-198.

[7] Wei W, Qin L, Jiang J . Image analysis and pattern recognition method of three-dimensional process in physical education teaching based on big data[J]...Journal of Electronic Imaging, 2022, 31:061811 - 061811.

[8] Abadi M, Richard B, Shamblen S, Drake C, Schweinhart A, Bokhour B. Achieving Whole Health: a Preliminary Study of TCMLH, a Group-Based Programme Promoting Self-Care and Empowerment Among Veterans:[J].Health Education & Behavior, 2022, 49(2):347-357.

[9] Gou C . An Integrated CoCoSo-CRITIC-Based Decision-Making Framework for Quality Evaluation of Innovation and Entrepreneurship Education in Vocational Colleges with Intuitionistic Fuzzy Information[J].Mathematical Problems in Engineering, 2022.

[10] Wang S . Application of Internet of Things Framework in Physical Education System[J].journal of internet technology, 2022, 23(2):307-320.

[11] Bo S L L . Promoting Effect of School Physical Education on Adolescent Mental Health Education--On the Cultivation of "Masculinity"[J].Psychology Research, 2022, 12(10):795-801.

[12] Taylor J, Avellone L, Brooke V, Wehman P, Inge K, Schall C. The impact of competitive integrated employment on economic, psychological, and physical health outcomes for individuals with intellectual and developmental disabilities[J].Journal of applied research in intellectual disabilities. JARID, 2022, 35(2):448-459.

[13] Qiu C, Xiao H, Wang L, Tian Y. Recent advances in integrated optical directed logic operations for high-performance optical computing: a review[J]. Frontiers in Optoelectronics: English Edition, 2022, 15(1):17.

[14] Liu Y, Sathishkumar V E, Manickam A . Augmented reality technology based on school physical education training[J].Computers and Electrical Engineering, 2022(99-):99.

[15] Liu Yong, Liu Yuanhai, Xiao Jun. Research on the application of Campus Running APP in the integration of university physical education inside and outside the classroom[J]. Education Modernisation, 2020, 7(4): 76-78.

[16] Robin B , Roger M , Lynnae J .Integrated Pest Management Data for Regulation, Research, and Education: Crop Profiles and Pest Management Strategic Plans[J].Journal of Integrated Pest Management, 2022(1):1.

[17] Johnston KC, Marttinen R . Examining evolutions of literacy integration with physical education and health in an after-school program [J].Journal of Adolescent & Adult Literacy, 2023.

[18] Zhu Z, Xu Z, Liu J . Flipped classroom supported by music combined with deep learning applied in physical education[J]. Applied Soft Computing, 2023.

[19] Zhenjun Zhu, Jienna Shang, Ping Wang, Ran Xiang, Dongqing Ye, Jian Zhou, Yan Shi, Zilu Qu. Research on the design and implementation of the teaching mode of "in-class and out-of-class integration" in the course of "University Physical Education" in Guangzhou University[J]. Journal of Guangzhou University (Natural Science Edition), 2022, 21(3): 89-95.

[20] Li Hongbin. The construction of "inside and outside class integration" overall curriculum system of physical education in Qinghai University [J]. Journal of Qinghai University (Natural Science Edition), 2010, 28(1): 92-97.

[21] Duan X X, Wang Y L, Dou W S, Kumar R, Saluja N. An Integrated Remote Control-Based Human-Robot Interface for Education Application[J].International journal of information technology and web engineering, 2022.

[22] Zha S, Jin Y, Wheeler R, Bosarge E. A mixed-method cluster analysis of physical computing and robotics integration in middle-grade math lesson plans [J].Computers & Education, 2022.

[23] Xie L, Han T, Zhou H, Zhang Z R, Han B, Tang A. Tuna Swarm Optimization: a Novel Swarm-Based Metaheuristic Algorithm for Global Optimization[J]. Hindawi Limited, 2021, 2021:1-22.

[24] Umarani O, Jyothi D . Deep Belief Networks Using Convolution Neural Networks Algorithm[J].

[25] Farahat M, Halavati R . Noise Robust Speech Recognition Using Deep Belief Networks[J]. Applications, 2016, 15(01):1650005.

Robust Image Tampering Detection and Ownership Authentication Using Zero-Watermarking and Siamese Neural Networks

Rodrigo Eduardo Arevalo-Ancona[1], Manuel Cedillo-Hernandez[2], Francisco Javier Garcia-Ugalde[3]
Instituto Politecnico Nacional, SEPI-ESIME Culhuacan, Mexico City, Mexico[1, 2]
Universidad Nacional Autonoma de Mexico, Facultad de Ingenieria, Mexico City, Mexico[3]

***Abstract*—The development of advanced image editing tools has significantly increased the manipulation of digital images, creating a pressing need for robust tamper detection and ownership authentication systems. This paper presents a method that combines zero-watermarking with Siamese neural networks to detect image tampering and verify ownership. The approach utilizes features from the Discrete Wavelet Transform (DWT) and employs two halftone images as watermarks: one representing the owner's portrait and the other corresponding to the protected image. A feature matrix is generated from the owner's portrait using the Siamese network and securely linked to the image's halftone watermark through an XOR operation. Additionally, data augmentation enhances the model's robustness, ensuring effective learning of image features even under geometric and signal processing distortions. Experimental results demonstrate high accuracy in recovering halftone images, enabling precise tamper detection and ownership verification across different datasets and image distortions (geometric and image processing distortions).**

***Keywords*—Zero-watermarking; tampering detection; ownership authentication; neural network**

I. Introduction

In recent years, digital image manipulation has become a growing concern with the development of advanced editing tools, many of which are easily accessible to the general public. This issue poses significant challenges to the integrity and authenticity of digital images in applications such as copyright protection, digital security, and forensic investigations [1], [2]. Manipulation techniques such as copy-move and splicing can compromise the content and image ownership [3], [4], [5]. Furthermore, ownership authentication is crucial to prevent unauthorized use or distribution, ensuring their protection [6].

Traditional watermarking methods verify the authenticity and ownership of images. However, these approaches degrade the image quality by embedding watermark signals into the visual content. To address this issue, zero-watermarking techniques preserve image quality by creating a master share that is not embedded into the image content. This method creates a master share by linking image features with a watermark for authentication, avoiding distortions and preserving image quality [7], [8].

Additionally, deep neural networks enhance tamper detection and image authentication by learning patterns and semantic representations from images. This approach significantly improves the efficiency of detecting manipulations and verifying image content.

This paper presents a robust zero-watermarking methodology that integrates zero-watermarking with Siamese neural networks for tamper detection and ownership verification. The approach utilizes the Discrete Wavelet Transform (DWT) to extract key features from images, which are then used to train the Siamese network and generate a feature matrix. This matrix is linked to the watermark—a proprietary halftone image—through an XOR operation, enabling accurate watermark recovery even in the presence of geometric distortions and signal processing attacks. Additionally, tamper detection is achieved by analyzing image blocks' eigenvalues to identify image manipulations effectively. The main contributions of this work are as follows:

- Enhance ownership authentication and tamper detection using a zero-watermarking technique that verifies image ownership and detects manipulations by combining image features extracted by the Siamese neural network with the watermark, ensuring security and integrity.
- Create a robust feature representation with Siamese Neural Networks. Robust feature representation is achieved because the neural network learns unique and invariant image features. One branch is trained with undistorted images, while the other is trained with distorted images, providing an efficient method improving the accuracy of ownership authentication and increases the method's robustness against manipulated images.
- Data augmentation techniques prevent overfitting and improve generalization across different types of images. This is essential because overfitting is common in neural network models, especially when working with limited or highly distorted data. In addition, data augmentation enhances feature extraction since the model learns more robust image features.
- The Siamese neural network learns and compares features, analyzing similarities and differences between the original halftone image's and distorted image's DWT coefficients. This is critical for tamper detection and authenticity verification.

- Optimized feature extraction using the Low-Low (LL) sub-band coefficients of the DWT. These filters remove irrelevant high-frequency details, focusing on the most relevant information for watermarking. This step is necessary to increase the robustness of the feature extraction process for watermark recovery, ensuring that important features are preserved even when the image has geometric or image processing distortions.
- A precise tamper detection method that compares the image blocks eigenvalues between the original halftone image and a distorted version, providing an additional layer of accuracy for tampering detection.

The paper is organized as follows: Section II reviews existing methods. Section III describes the proposed technique. Section IV presents and analyzes experimental results. Finally, Section V concludes with the study's advantages and limitations.

II. Literature Review

This section presents an overview of recent techniques developed for image tampering detection and image authentication, providing the necessary context for the method proposed in this paper.

Several approaches are focused on detecting tampered regions by extracting specific image features. For example, Xing et al. [9] used a high-pass filter to capture edge information, concentrating on modified regions to identify manipulated areas. Alsughayer et al. [10] employed a U-Net architecture to extract residual noise from images, identifying tampered regions in remote sensing data. However, noise-based methods are susceptible to false positives when noise is introduced by natural image compression or processing.

With the rise of deep learning, some network-based solutions for tamper detection were developed, such as Priyadharsini et al. [11], who applied a modified GoogleNet model to extract features using a nearest neighbor algorithm for splicing detection. However, this method is computationally expensive and is prone to overfitting with small datasets. Ren et al. [12] introduced a neural network for low-level feature extraction, focusing on geometric details and semantic segmentation for copy-move detection. Goel et al. [13] used a dual neural network for inherent feature extraction, although their method requires large amounts of labeled data for accurate detection. Hosny et al. [14] proposed a method based on deep neural networks for extracting specific image features, while Chu [15] used a generative neural network to detect forged regions. Das et al. [16] leveraged MobileNet V2 to detect manipulated areas, highlighting the potential of lightweight models for real-time applications. Nikalie et al. [17] combined convolutional neural networks with Local Binary Pattern (LBP) analysis to detect texture inconsistencies. Mallick et al. [18] incorporated VGG16 and VGG19 models to identify manipulations across images with different compression levels, but their approach requires significant computational resources. Dai et al. [19] employed the Xception model to extract image edges and textures, offering improvements in detecting fine manipulations.

On the other hand, recent zero-watermarking methods for image authentication integrated neural networks to improve robustness. Xiang et al. [20] combined watermarking with features generated by a ResNet-based neural network, which relied on correlations between neural network layer responses to verify authenticity. However, these models are susceptible to geometric distortions. Dong et al. [21] proposed a method using NasNet-Mobile features and Discrete Cosine Transform (DCT) coefficients to create a master share, allowing watermark recovery without degrading the image. Similarly, Li et al. [22] introduced an image authentication method based on ConvNeXt layers and Swin Transformer optimization for feature extraction, although the approach lacks comprehensive evaluation against geometric attacks. He et al. [23] further advanced the field by integrating convolutional layers based on the Swin Transformer to optimize feature extraction for zero-watermarking, yet the robustness under signal processing distortions remains underexplored.

While significant progress has been made in both neural network-based tamper detection and watermarking techniques, these methods still need to improve their robustness, especially under geometric distortions and signal processing attacks. In contrast, the method proposed in this paper addresses these limitations for image ownership authentication and detects image tampering by combining Siamese neural networks with zero-watermarking. This ensures high accuracy in tamper detection and strong resilience to various distortions without compromising image quality. Additionally, the Discrete Wavelet Transform (DWT) allows for more efficient feature extraction, focusing on relevant low-frequency components, further enhancing robustness.

III. Proposed Image Tampering Detection and Ownership Authentication Method

The proposed algorithm introduces a robust zero-watermarking technique for image protection, creating a master share, which is generated by logically linking the halftone image representation with features extracted from the watermark, represented by the owner's portrait, through a neural network (Fig. 1(a)). The preprocessing stage begins by dividing the input image into 32x32 pixel blocks and applying the Discrete Wavelet Transform (DWT) to extract the Low-Low (LL) sub-band coefficients, which enhance the system's robustness against geometric distortions and signal processing attacks.

The features extracted by the Siamese neural network are carefully selected because they capture invariant and distinctive representations of the watermark, which are crucial for an accurate ownership verification and tamper detection. The Floyd-Steinberg dithering algorithm generates the halftone effect [24], [25].

For the tamper detection and ownership verification stage, the master share is combined with the features extracted by the Siamese network to recover the halftone representation of the protected image. The eigenvalues from the feature matrices of the recovered and potentially manipulated images are compared to identify image forgery regions. (Fig. 1(b)).

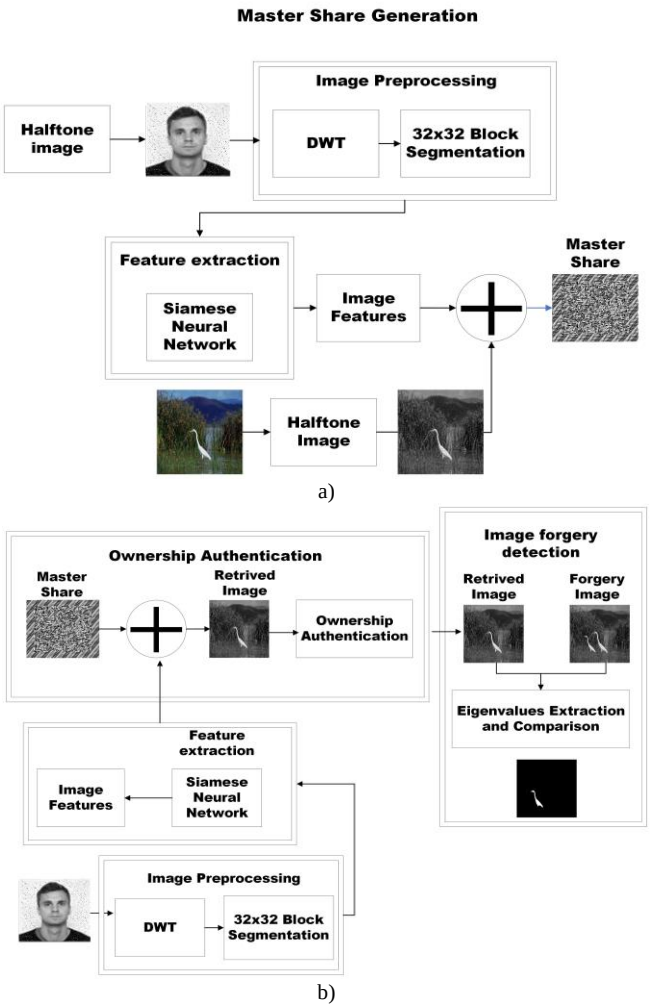


Fig. 1. (a) General diagram of the proposed master share generation algorithm (b) Ownership authentication and tamper detection procedures.

The proposed method consists of the following stages:

Preprocessing: Halftone representations of both the image and the watermark are created. The halftone watermark is decomposed using the Haar wavelet transform (DWT), focusing on the LL sub-band. This sub-band is divided into 32x32 blocks used to train the neural network. Dividing the image into smaller blocks identifies specific features from each image region more effectively, improving the accuracy of feature extraction and reducing computational complexity.

Feature extraction for the master share generation: The Siamese neural network extracts key features from the watermark. These features are combined with the image's halftone representation to generate the master share.

Image ownership authentication: The halftone image is retrieved by linking the watermark's extracted features with the master share. This step validates image ownership, offering robust intellectual property protection.

Image tampering detection: The eigenvalues from the halftone and retrieved images are compared to detect forgeries. Eigenvalues are sensitive to structural changes in the image, making them an effective tool to identify image tampering.

A. Preprocessing

In the preprocessing stage, the halftone image representations are generated for the image and the owner's portrait (watermark). The halftone representation simplifies the visual content with binary patterns, eliminating unnecessary details. Then, the owner's halftone image is transformed into the frequency domain using the DWT with Haar wavelets. Haar wavelets are effective for image processing tasks capturing spatial and frequency information. The DWT decomposes the image into frequency sub-bands, each capturing specific image characteristics. The low-low (LL) sub-band is selected for further processing and contains low frequency image components. These components represent the most important structural details, such as edges and large-scale patterns, which are less susceptible to distortion like noise or compression and eliminate redundant features. The LL sub-band coefficients are divided into 32x32 pixel blocks to optimize the system's robustness. This segmentation enhanced feature extraction because each block represents an image region, allowing the neural network to learn specific features. In addition, when working with smaller blocks, the overall processing time is reduced, and memory usage is optimized, making the algorithm more scalable and suitable for real-time applications, highlighting the system's adaptability.

B. Feature Extraction for the Master Share Generation

In this stage, the Siamese neural network extracts relevant features from the halftone watermark. The Siamese network is designed to compare two images and learn invariant features robust to distortions and manipulations (Fig. 2). This neural network model consists of two identical subnetworks, ensuring the feature matrix contains specific representations from the watermark. In this case, one subnetwork processes the original DWT watermark blocks while the other processes a distorted version of the watermark DWT blocks.

The features extracted from the watermark by the Siamese network are used to construct a feature matrix, which encodes the unique characteristics of the watermark. This feature matrix contains patterns from the watermark and the distorted watermark version with geometric and image processing attacks. Then, the feature matrix is combined with the halftone representation of the image. The combination is achieved using an XOR operation.

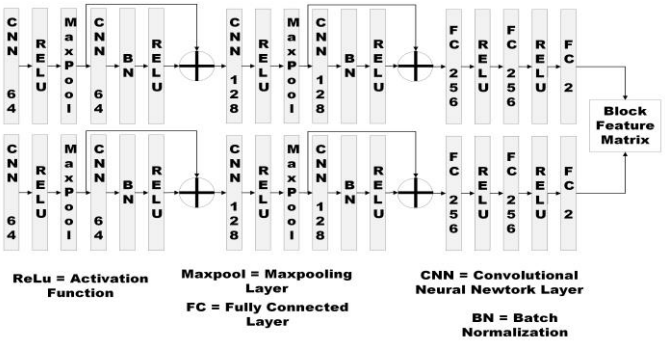


Fig. 2. Siamese neural network architecture.

The resulting master share contains a visual structure that encrypts the image and the semantic features of the watermark. This technique preserves the image quality since no watermark is embedded directly into the visual content of the image, maintaining its integrity.

The parameters and configurations used in the training process to increase the robustness and accuracy in the system are presented in Table I. The Siamese neural network architecture consists of two identical branches. To train the neural network one branch processes the DWT watermark blocks, and the second branch processes the DWT distorted watermark blocks. This architecture is especially useful when the data is limited. The purpose of the network is to compare image features from both input images and identify unique patterns in each region. The network creates a feature matrix (f_{vm}) (2) based on the vectors generated from the output of the Siamese neural network (v_1, v_2) related to each image region, representing the unique characteristics.

$$v_1 = f_\theta(I_1)$$
$$v_2 = f_\theta(I_2) \quad (1)$$

$$f_{vm} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad (2)$$

Then feature matrix with the Siamese neural network output (f_{vm}) has a size of (3) and is restructured to the image halftone size (f_m) (4).

$$size = (2, (\frac{m\ x\ n}{block\ size})) \quad (3)$$

$$fm = \begin{bmatrix} c_{1,1} & \cdots & c_{1,n} \\ \vdots & \ddots & \vdots \\ c_{m,1} & \cdots & c_{m,n} \end{bmatrix} \quad (4)$$

m is the height of the image and *n* the width of the image.

1) Neural network training: One key aspect is data augmentation, which enhances the model's feature detection. Data augmentation is important when the amount of data is limited, as it artificially increases the dataset by applying distortions to the input images. The image processing attacks applied during augmentation include different image filters for image blurring (average, median, Gaussian, motion blurring), JPEG compressions (quality factor = 90, 60, 10), noise addition (Gaussian noise σ = 0.09, 0.009 and salt and pepper), rotations (random rotations from 10° to 350°), translations (image pixels shift in x, y and xy axis = 50, 100, 150), cropping (random cropping with sizes in x and y= 50, 100, 150), and scaling (scaling factors = 2, 0.5, 0.25).

The application of these augmentation techniques ensures that the network learns robust, invariant features to obtain high ownership authentication and image tampering detection accuracy even in the presence of noise, distortions, or other forms of manipulation.

TABLE I. SIAMESE NEURAL NETWORK PARAMETERS

Epochs	Learning rate	Momentum	Optimazer	Batch size
3	0.0001	0.9	Stochastic Gradient Descent	64

2) Master share generation: The features extracted from the watermark by the Siamese network are used to construct a f_m, which encodes the unique characteristics of the watermark (W).

This feature matrix is then combined with the halftone representation of the watermark using an XOR operation ($\oplus$) to create the master share (*MS*) (5) (see Fig. 3). This operation effectively encrypts the image while preserving its quality, as no watermark is embedded directly into the visual content of the image. This XOR-based encryption method securely. The resulting master share (see Fig. 4(d)) is then stored and distributed. To recover the original image, the master share must be used together with the watermark.

$$ms = f_m \oplus H \quad (5)$$

Once the master share has been generated and stored, image ownership authentication can be carried out. The following sections detail how this ownership authentication process works and how the master share is utilized for both authentication and tamper detection.

C. Image Ownership Authentication and Tamper Detection

To verify ownership, the master share used watermark during recovery even if it is distorted. This element ensures that only the owner, who possesses the watermark, can authenticate the image. The authentication process involves recovering the halftone image from the master share using both the owner's portrait and the master share. This procedure, illustrated in Fig. 5, confirms the image's ownership. The ownership is confirmed by reconstructing the halftone image and verifying its authenticity.

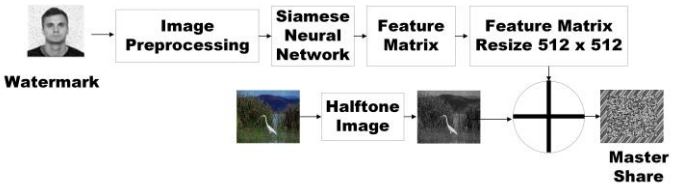


Fig. 3. Master share generation process.

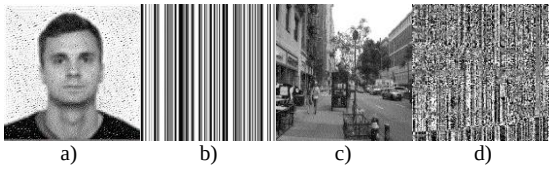


Fig. 4. a) Halftone watermark, b) Feature matrix generated from the Siamese neural network, c) Halftone image, d) Master share.

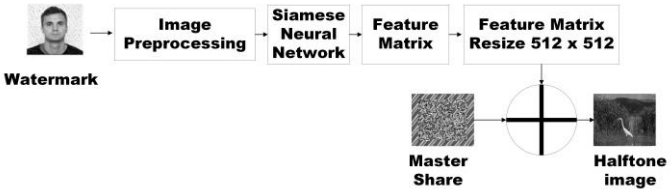


Fig. 5. Image halftone recovery process.

This process involves decoding the master share (*ms*) to reconstruct the halftone image (H_r) by combining the feature matrix from the Siamese neural network (f_m) with the master share using the XOR ($\oplus$) logical operation (6).

$$H_r = f_m \oplus ms \tag{6}$$

Once ownership has been confirmed through the successful reconstruction of the halftone image, it is important to ensure that the image has not been tampered. While the ownership authentication process guarantees that only the owner can verify the image, the tampering detection process compares the original and recovered halftone images to detect unauthorized modifications. The following section details how the system uses eigenvalues to detect manipulations by comparing the structural integrity of the original and recovered images, ensuring that the image remains unchanged and free from tampering.

D. Tampering Detection

The manipulation detection used the recovered and manipulated halftone images to identify any potential alterations. In this stage, eigenvalues are calculated for both the recovered halftone and potentially manipulated images. The images are divided into smaller blocks, and the eigenvalues for each block are calculated. These eigenvalues represent distinct structural characteristics of each image and are compared to identify discrepancies. The eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ are computed by (7).

$$v_I = \det(I - \lambda) \text{ and } v_m = \det(I_m - \lambda) \tag{7}$$

where I is the retrieved halftone image, I_m is the manipulated image, v_I is the vector with the eigenvalues from the retrieved image and $v_m$is the vector with the eigenvalues from the manipulated image. The eigenvalues comparison is realized by the Euclidean distance (*d*) (8).

$$di = \sqrt{(\lambda_I - \lambda_m)^2} \tag{8}$$

For each block, if the distance exceeds a predefined threshold (*th*), the block (b_d) is considered manipulated (9).

$$b_d = \begin{cases} 0 & if \quad d \le th \\ 1 & if \quad d > th \end{cases} \tag{9}$$

The results of this process, illustrating the detection of manipulations, are shown in Fig. 6. This method enhances the precision of manipulation detection. Block-based eigenvalue comparisons allow localized forgery detection, ensuring that even subtle tampering can be identified, thus offering a more detailed and accurate analysis of image manipulations.

a) b) c)

Fig. 6. a) Original image, b) Tampered image, c) Forgery detection.

In the following section, the experimental results obtained from this methodology demonstrating its effectiveness for ownership authentication and image tampering detection.

IV. Experimental Results

This section presents the experimental results of the proposed algorithm for both ownership authentication and image tampering detection. The image watermark was subjected to a wide range of image processing techniques and geometric distortions, simulating real scenarios to assess the method's robustness. The experimental results demonstrate the algorithm's effectiveness in accurately verifying ownership and detecting manipulations, even when the watermark has been significantly altered.

The algorithm was implemented in a system equipped with an NVIDIA GTX 960 graphics card and an Intel Core i7-6700 processor running at 3.4 GHz. This hardware configuration provided a suitable environment for efficiently executing the algorithm within the PyTorch Python framework, ensuring smooth performance during training and testing. A 512 x 512 pixel halftone representation of the author's face was used as the watermark for the experiments. The proposed algorithm's performance was evaluated using three widely used image datasets, showcasing its ability to handle diverse scenarios and conditions:

MICC-F220: This dataset contains 220 RGB images, equally divided into 110 manipulated and 110 non-manipulated images. It evaluates owner authentication and tamper detection accuracy in a balanced setting [26].

MICC-F2000: This dataset is comprised of 2,000 RGB images (700 manipulated and 1,300 non-manipulated), which is used for a more extensive assessment of the algorithm's performance across diverse scenarios and conditions [26].

CASIA V2: This larger dataset includes 12,613 RGB images, 5,123 manipulated, and 7,490 non-manipulated images. Due to its size and diversity, CASIA V2 provides a robust evaluation framework, testing the algorithm's scalability and effectiveness across a wide range of image manipulations [27], [28].

These datasets cover different image manipulation scenarios to evaluate the algorithm's robustness and accuracy. The experiments using these datasets quantitatively assess the algorithm's ability to recover the halftone images for ownership authentication and image forgery detection. The results offer valuable insights into the algorithm's effectiveness in different conditions, showcasing its ability to detect subtle manipulations and authenticate image ownership.

A. Assessment of Image Ownership Authentication

The proposed method's performance is assessed in terms of ownership authentication, emphasizing its robustness against different image processing and geometric distortions. To ensure this robustness, two metrics are employed: Bit Error Rate (BER) and Normalized Cross-Correlation (NC). These metrics determine the algorithm's performance under image processing and geometric distortions applied to the watermark.

The BER measures the error bits between the original and recovered halftone images. A lower BER indicates that the algorithm successfully retrieves the watermark (10).

$$BER = \frac{Total\ Incorrect\ pixels}{Total\ pixels} \tag{10}$$

The NC metric measures the similarity between the original and recovered halftone images. The NC value ranges from 0 to 1, where a value of 1 indicates similarity between the images, and a value of 0 indicates no similarity (11).

$$NC = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(W(i,j)W_r(i,j))}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(W(i,j))^2}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(W_r(i,j))^2}} \tag{11}$$

where, $H(i,j)$ is the halftone image, H_r is the retrieved halftone image M and N are the dimensions of the images. These metrics ensure a comprehensive evaluation of the algorithm's watermark recovery stage performance for ownership authentication, even under image distortion and manipulation.

TABLE II. HALFTONE IMAGE RECOVERY WITH WATERMARK GEOMETRIC DISTORTIONS

	Rotation 80 °	Rotation with cropping 250 °	Translation X = 90 Y = 90	Translation X = 160
Distorted watermark				
Retrieved halftone image				
	NC = 0.999 BER = 0	NC = 0.999 BER = 0	NC = 0.999 BER = 0	NC = 0.999 BER = 0
	Cropping upper left	Center cropping	Scale 64 x 64	Scale 1064 x 1064
Distorted watermark				
Retrieved halftone image				
	NC = 0.999 BER = 0	NC = 0.999 BER = 0	NC = 0.999 BER = 0	NC = 0.999 BER = 0

Tables II and III demonstrate the proposed method's robustness in recovering the halftone image even when the watermark has geometric distortions. Under distortions like rotation without cropping, translation, and scaling, the recovery performance remains nearly perfect, with BER = 0 and NC = 0.9999. Minor recovery errors occur with rotations involving cropping and downscaling, where BER reaches 0.0078 and NC drops slightly to 0.9914. These results demonstrated the method's robustness against geometric distortions, ensuring an accurate watermark recovery for ownership verification.

TABLE III. HALFTONE IMAGE RECOVERY ASSESSMENT WITH WATERMARK GEOMETRIC DISTORTIONS

Distortion	BER	NC	Distortion	BER	NC
No attack	0	0.9999	Translation x = 80 pixels	0	0.9999
Rotation 15° with no cropping	0	0.9999	Translation x = 150 pixels	0	0.9999
Rotation 55° with no cropping	0	0.9999	Translation x = 20, y = 20 pixels	0	0.9999
Rotation 110°	0	0.9999	Translation x = 80, y= 80 pixels	0	0.9999
Rotation 135° with no cropping	0	0.9999	Translation x = 150, y = 150 pixels	0	0.9999
Rotation 20° with cropping	0.0078	0.9914	Upper left cropping 100 x 100	0	0.9999
Rotation 100° with cropping	0.0078	0.9923	Upper right cropping 100 x 100	0	0.9999
Rotation 265° with cropping	0.0078	0.9901	Botton right cropping 100 x 100	0	0.9999
Rotation 320° with cropping	0.0078	0.9914	Center cropping 100 x 100	0	0.9999
Translation y = 20 pixels	0	0.9999	Scale 256 x 256	0	0.9999
Translation y = 80 pixels	0	0.9999	Scale 64 x 64	0.0078	0.9914
Translation y = 150 pixels	0	0.9999	Scale 640 x 640	0	0.9999
Translation x = 20 pixels	0	0.9999	Affine transform $\begin{bmatrix} 5 & 10 & 0 \\ 10 & 5 & 0 \end{bmatrix}$	0	0.9999

Tables IV and V illustrate the effectiveness of the proposed method for recovering halftone images when the watermark is modified with image processing distortions, such as JPEG compression, affine transformations, histogram equalization, and noise addition. In most cases, the NC is 0.999, and the BER is 0, indicating high accuracy. Even under more challenging conditions, such as JPEG compression at a quality factor of 30, the method still maintains a high accuracy with NC = 0.992 and BER = 0.007. The proposed technique remains robust against distortions like salt-and-pepper noise and Gaussian filtering, achieving NC = 0.999 and BER = 0,

demonstrating its robustness in the halftone image recovery stage for ownership authentication.

TABLE IV. HALFTONE IMAGE RECOVERY ASSESSMENT WITH WATERMARK IMAGE PROCESSING DISTORTIONS

Distortion	BER	NC	Distortion	BER	NC
JPEG Quality Factor = 70	0	0.9999	Gaussian Filter kernel = 7 x 7	0	0.9999
JPEG Quality Factor = 50	0	0.9999	Median filter kernel = 7 x 7	0.0078	0.9916
JPEG Quality Factor = 30	0	0.9999	Gaussian noise $\mu = 0, \sigma = 0.009$	0	0.9999
Blurring kernel = 5 x 5	0.0078	0.9920	Gaussian noise $\mu = 0, \sigma = 0.09$	0	0.9999
Gaussian Filter kernel = 5 x 5	0.0078	0.9914	Salt and pepper noise 0.005	0.0078	0.9915
Median filter kernel = 5 x 5	0.0078	0.9911	Salt and pepper noise 0.05	0.0078	0.9914
Average filter kernel = 5 x 5	0.0078	0.9910	Gamma correction $\gamma = 1.25$	0.0078	0.9914
Blurring kernel = 7 x 7	0	0.9999	Gamma correction $\gamma = 0.80$	0.0078	0.9919
Histogram equalization	0	0.9999	Bright adjust	0.0078	0.9916

TABLE V. HALFTONE IMAGE RECOVERY WITH WATERMARK IMAGE PROCESSING DISTORTIONS

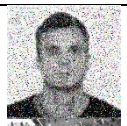
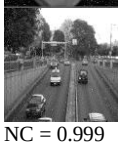
	JPEG 70	JPEG 30	Affine Transform	Histogram
Distorted watermark				
Retrieved halftone image				
	NC = 0.999 BER = 0	NC = 0.992 BER = 0.007	NC = 0.999 BER = 0	NC = 0.999 BER = 0
	Gamma 1.5	Salt and pepper noise 0.09	Gaussian filter kernel = 7 x 7	Blurring kernel = 5x5
Distorted watermark				
Retrieved halftone image				
	NC = 0.999 BER = 0	NC = 0.999 BER = 0	NC = 0.999 BER = 0	NC = 0.999 BER = 0

TABLE VI. HALFTONE IMAGE RECOVERY ASSESSMENT WITH WATERMARK COMBINED DISTORTIONS

Distortion	BER	NC	Distortion	BER	NC
$\gamma = 1.25$ and salt and pepper noise 0.005	0	0.9999	JPEG Quality Factor=30 and Blurring kernel=5x5	0	0.9999
$\gamma = 1.8$ and salt and pepper noise 0.005	0	0.9999	Bright adjust and scaling 1024x1024	0	0.9999
Rotation 35° and bright adjust	0	0.9999	JPEG Quality Factor=30 and Scaling 64x64	0	0.9999

Table VI evaluates the halftone image recovery under combined distortions applied to the watermark.

For all tested combinations, including gamma correction with salt-and-pepper noise, JPEG compression with blurring, brightness adjustment with scaling, and rotation with brightness adjustment, the method achieves NC = 0.9999 and BER = 0. These results indicate that the recovery process is highly robust even when multiple distortions are applied simultaneously to the watermark, ensuring the integrity and authenticity of the halftone image.

B. Tampering Detection Assessment

Following the evaluation of image ownership authentication, assessing the algorithm's effectiveness in detecting image tampering is essential, as both tasks are closely related. The tampering detection accuracy relies on the halftone image's successful recovery during the ownership authentication stage.

Table VII presents the results, demonstrating the robustness of the proposed method in detecting image forgery by analyzing the recovered and manipulated halftone images. These results highlight the system's effectiveness by accurately detecting any image manipulations. The proposed image tampering detection and localization algorithm is evaluated using five key metrics: accuracy, precision, recall, F1 score, and mean squared error (MSE).

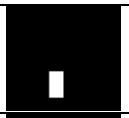
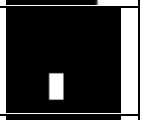
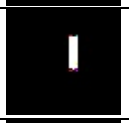
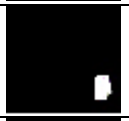
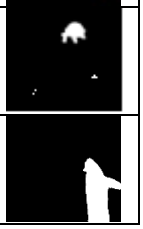
The accuracy represents the proportion of correctly detected pixels relative to the total number of pixels, indicating how well the algorithm identifies manipulated areas (12).

$$acc = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{12}$$

where T_p = true positives, T_n = true negatives, F_p = false positives, F_n = False Negatives. The precision measures the ratio of correctly detected pixels from the manipulated region (13).

$$precision = \frac{Tp}{Tp + Fp} \tag{13}$$

TABLE VII. IMAGE FORGERY DETECTION

Database	Original Image	Tampered Image	Ground Truth	Tampering Detection
MIC-F220				
				
MIC-F2000				
				
CASIA V2				
				

The recall measures the proportion of correctly detected pixels in the manipulated region (14).

$$recall = \frac{Tp}{Tp + Fn} \tag{14}$$

The F1 score provides a measure of the algorithm's performance. A high F1 score indicates that the algorithm achieves minimize false positives and false negatives (15)

$$F1 = \frac{2x(Precision\ x\ Recall)}{Precision + Recall} \tag{15}$$

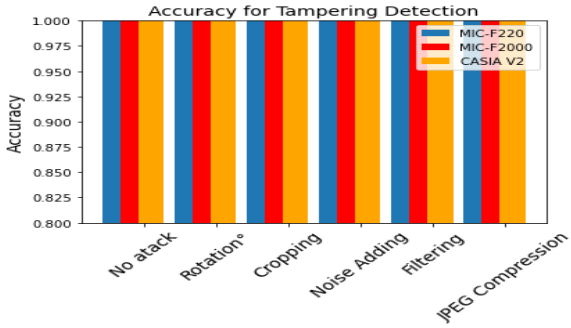


Fig. 7. Accuracy for watermark tampering with geometric distortions and image processing distortions.

Fig. 7 to Fig. 10 provide a detailed evaluation of the proposed method's effectiveness in detecting image manipulations. The results show high efficiency in detecting tampered areas, even when the watermark has been distorted. However, some loss of features is observed in the halftone image recovery when image processing distortions are applied to the watermark, which introduces errors in the recovery process. This error can impact the accuracy of tamper detection. Despite these challenges, the method identifies manipulations efficiently. High accuracy ensures that most manipulations are correctly identified, while precision and recall measure the method's ability to distinguish tampered areas from untampered ones. The F1 score, a balance between precision and recall, ensures a high performance even when the image is distorted.

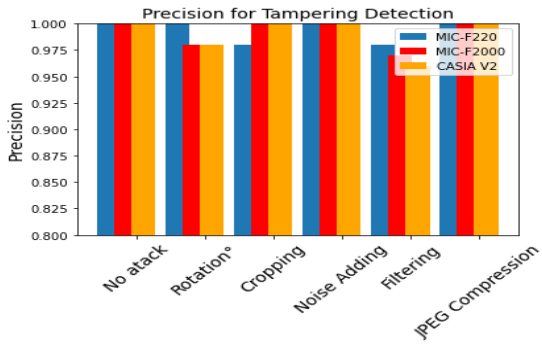


Fig. 8. Precision for watermark tampering with geometric distortions and image processing distortions.

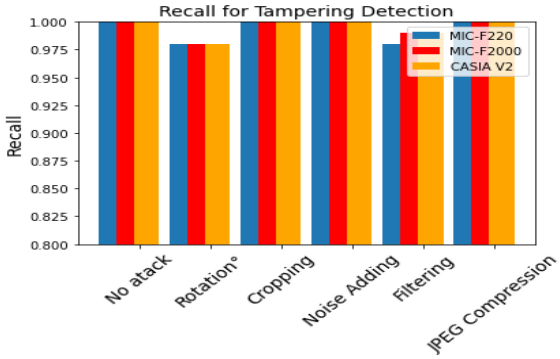


Fig. 9. Recall for for watermark tampering with geometric distortions and image processing distortions.

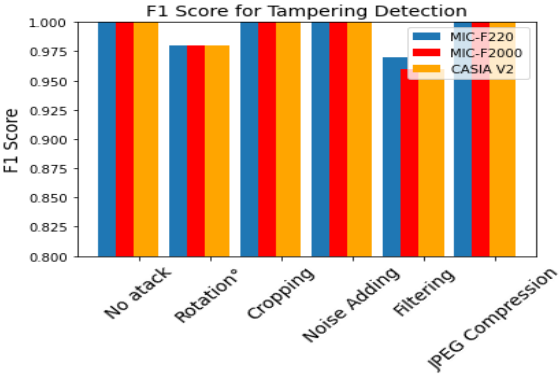


Fig. 10. F1 score for watermark tampering with geometric distortions and image processing distortions.

C. Ownership Authentication Performance Comparison

In this section, the effectiveness of the proposed image ownership authentication algorithm is compared to existing methodologies. This comparison highlights the proposed method's robustness against different image distortions relative to existing solutions in terms of accuracy, robustness, and efficiency.

Table VIII compares the proposed method with other existing techniques. One of the main advantages of the proposed method is block segmentation and the neural network model feature extraction for each region of the image. This processing provides the neural network with recognition of specific patterns related to each block, which generates a robust feature matrix. This methodology allows for specific feature analysis to improve the algorithm's capacity in image authentication. The comparison depicted that the method is more robust against most distortions; however, image recovery can generate errors in the case of filtration distortions. The results from the zero-watermarking comparison show that the proposed method outperforms existing approaches. This method uses LL coefficients from the DWT to train a Siamese network, generating a robust feature matrix. This generates higher accuracy in the recovery halftone stage. The results from Table VIII indicate that the proposed method is highly resistant to image quality loss, crucial for scenarios involving image storage or transmission. Also, the present technique shows robustness to rotations, scaling, and translations, while other methods obtain a BER of 0.02 for rotations of only 10°.

D. Image Forgery Detection Comparison

Table IX presents a comparison between the proposed method and other techniques developed for tamper detection. The results demonstrate the effectiveness of the proposed method in detecting and localizing image tampering across image databases. Furthermore, the results show that the proposed method achieves higher efficiency than other approaches, effectively identifying whether an image has been manipulated and accurately pinpointing the manipulated areas.

TABLE VIII. ZERO-WATERMARKING COMPARISON

	Xiang et al. [20]	Dong et al. [21]	Li et al. [22]	He et al. [23]	Proposed method
Methodology	Correlation between shallow and deep features. Use the ResNet model as a feature extractor.	DCT low frequencies from the features of the NasNet-Mobile model.	ConvNext blocks from the ZWnet	Redundant feature shrinkage and removal (SRFENet)	LL-DWT coefficients to train a Siamese Network and create a feature matrix from each image region.
Image type	Medical images	Medical images	Natural images	Natural images	Natural images
Watermark size	256 x 256	---	---	512 x 512	512 x 512
Gaussian Noise	σ = --- NC = 0.9836	σ = --- NC = 0.93	σ = 0.005 NC = 1	σ = 0.010 BER = 0.001	σ = 0.09 NC=0.9999 BER=0
Salt and Pepper noise	NC = 0.9836	---	Factor = 0.01 NC = 0.9922	Factor = 0.01 BER = 0.01	Factor = 0.05 NC=0.9999 BER=0
JPEG	Quality Factor = 10 NC = 0.9834	Quality Factor = 75 NC = 1	---	Quality Factor = 30 BER= .01	Quality Factor = 30 NC=0.9999 BER=0
Gaussian Filter	0.9834	---	NC = 1	BER = 0.015	NC=0.9920 BER=0 0078
Median Filter	Kernel 3x3 NC = 0.9833	Kernel 5x5 NC = 0.92	---	Kernel 3x3 BER= .01 Kernel 5x5 BER = 0.015	Kernel 3x3 NC=0.9920 BER=0.0078
Rotation	20° NC = 0.9835	---	15° NC = 0.9688	10° BER = 0.02	135° NC=0.9999 BER=0
Cropping	Size = 1/3 NC = 0.9834	Size = 10 % 0.94	Size = 1/8 NC = 0.9063	---	Size = 100 x 100 NC=0.9999 BER=0
Scaling	Factor 0.8 NC = 0.9835	Factor 0.3 NC = 0.88	---	Factor 0.8 BER = 0.01	Size = 64 x 64 NC=0.9999 BER=0
Translation	---	X = 15% NC = 0.94 Y = 10% NC = 0.86	---	---	X = 100, Y = 100 NC=0.9999 BER=0

TABLE IX. IMAGE FORGERY DETECTION COMPARISON

	Das et al [16]	Nikalje et al. [17]	Mallick et al. [18]	Dai et al. [19]	Proposed method
Tampering detection	Splicng	Splicing and copy-move	Splicing and copy-move	Splicing and copy-move	Splicing and copy-move
Technique	CNN based on transfer learning MobileNet	CNN and Local Binary Pattern	Pretrained VGG16 and VGG19	Dual-Net DeepLab V3	Siamese Neural Network and DWT LL coefficients
Accuracy	0.9301	0.9901	VGG16 = 0.944 VGG19 = 0.995	0.8725	MIC-F220 = 0.993 MIC-F2000 = 0.991 CASIA V2 = 0.997
Precision	0.926	0.9581	---	---	MIC-F220 = 0.986 MIC-F2000 =0.983 CASIA V2 =0.985
Recall	0.966	0.9661	---	---	MIC-F220 = 0.994 MIC-F2000 = 0.997 CASIA V2 = 0.985

V. CONCLUSION

The method proposed in this paper provides a robust solution for owner authentication and image manipulation detection. The use of halftone images for image authentication based on the Siamese neural network features implemented in the zero-watermarking technique increased the efficiency of the watermark recovery from the master share. Furthermore, this technique provides an additional security stage by encrypting the image in the master share. In addition, the results show robustness in the image halftone recovery process when the watermark is distorted with geometric and image processing attacks. This is reflected in a low error and high similarity between the recovered and original halftone images. The coefficients belonging to the LL sub-band of the DWT and their segmentation into blocks allow the neural network to recognize unique patterns from each image region even though the watermark is distorted. On the other hand, the architecture of two branches from the Siamese neural network detects unique and invariant characteristics related to the watermark.

However, the generated error in the image recovery process increases when image processing distortions are applied to the watermark because some image features are deleted; despite this, a low error can be observed in the halftone image recovery. Furthermore, the recovered halftone image process enhances the effectiveness of the proposed image tampering detection method. The retrieved halftone image with minimal error serves as a reference for detecting discrepancies between the original and potentially manipulated images. The comparison with existing methods highlights the efficiency of forgery detection. The proposed methodology's main contribution focuses on its double function since it performs image owner authentication without distorting the image and effectively detects image tampering. In addition, the retrieved image analysis from the owner authentication process allows the detection and localization of image tampering.

ACKNOWLEDGMENT

This work was supported in part by the Instituto Politecnico Nacional (IPN), Universidad Nacional Autonoma de Mexico (UNAM) under the Direccion General de Asuntos del Personal Academico (DGAPA) research project under grant PAPIITIT100123, and in part by Consejo Nacional de Humanidades, Ciencias y Tecnologias (CONAHCYT).

REFERENCES

[1] S. Tyagi, D. Yadav, “A detailed analysis of image and video forgery detection techniques”. Vis Comput vol. 39, pp. 813–833, 2023, doi; https://doi.org/10.1007/s00371-021-02347-4.

[2] K. B. Meena and V. Tyagi, “Image Forgery Detection: Survey and Future Directions”, Data Engineering and Applications, vol. 2, pp. 163-194, 2019, doi: https://doi.org/.

[3] M. Maashi, H. Alamro, H. Mohsen, N. Negm, G. P. Mohammed, N. A. Ahmed, S. S. Ibrahim and M. I. Alsaid, “Modeling of Reptile Search Algorithm With Deep Learning Approach for Copy Move Image Forgery Detection”, IEEE Access, vol. 11, pp. 87297 - 87304, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3304237.

[4] T. Pomari, G. Ruppert, E. Rezende, A. Rocha and T. Carvalho, “Image Splicing Detection Through Illumination Inconsistencies and Deep Learning”, IEEE International Conference on Image Processing, Athens, Greece, 2018, doi: https://doi.org/10.1109/ICIP.2018.8451227.

[5] Chen, H., Han, Q., Li, Q. et al., “Digital image manipulation detection with weak feature stream”. Vis Comput, vol. 38, pp. 2675–268, 2022, doi: https://doi.org/10.1007/s00371-021-02146-x.

[6] R. Sinhal and I. A. Ansari, “Machine learning based multipurpose medical image watermarking”, Neural Computing and Applications, vol. 35, 2023, doi: https://doi.org/10.1007/s00521-023-08457-5.

[7] Y. Li, J. Li, U. A. Bhatti, J. Ma, D. Li and F. Dong, “Robust Zero-watermarking Algorithm for Medical Images Based on ORB and DCT”, ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter), Taiyuan, China, 2023, doi: https://doi.org/10.1109/SNPD-Winter57765.2023.10223992.

[8] G. Sun, J. Li, U. A. Bhatti, J. Ma, F. Dong and Y. Li, “Robust Zero-Watermarking Algorithm for Medical Images Based on AGAST-LATCH and DCT”, ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter), Taiyuan, China, 2023, doi: https://doi.org/10.1109/SNPD-Winter57765.2023.10224055.

[9] J. Xing, X. Tian and Y. Han, “A Dual-channel Augmented Attentive Dense-convolutional Network for power image splicing tamper detection”, Neural Computing and Applications, 2024, doi: https://doi.org/10.1007/s00521-024-09511-6.

[10] R. Alsughayer, M. Hussain, F. Saeed and H. AboalSamh, “Detection and localization of splicing on remote sensing images using image-to-image transformation”, Applied intelligence, vol. 53, pp. 13275–13292, 2023, doi: https://doi.org/10.1007/s10489-022-04126-7.

[11] S. Priyadharsini and K. K. Devi, “Effective image splicing detection using deep neural network, International Journal of Wavelets”, Multiresolution, vol. 21, nº 2, pp. 2250051-2250079, 2022, doi: https://doi.org/10.1142/S0219691322500515.

[12] R. Ren, S. Niu, J. Jin, K. Xiong and H. Ren, “ERINet: efficient and robust identification network for image copy-move forgery detection and localization”, Applied Intelligence, vol. 53, p. 16170–16191, 2023, doi: https://doi.org/10.1007/s10489-022-04104-z.

[13] N. Goel, S. Kaur and R. Bala, “Dual branch convolutional neural network for copy move forgery detection”, IET Image Processing, 2020, doi: https://doi.org/10.1049/ipr2.12051.

[14] K. M. Hosny, A. M. Mortda, M. M. Fouda and N. A. Lashin, “An Efficient CNN Model to Detect Copy-Move Image Forgery”, IEEE Access, pp. 48622 - 48632, 2022, doi: https://doi.org/10.1109/ACCESS.2022.3172273.

[15] L. Chu, “Research on Image Tampering Detection and Localization Based on Iterative GAN”, IEEE International Conference on Sensors, Electronics and Computer Engineering (ICSECE), Jinzhou, China, 2023, doi: https://doi.org/ 10.1109/ICSECE58870.2023.10263546.

[16] D. Das and R. Naskar, “Image Splicing Detection based on Deep Convolutional Neural Network and Transfer Learning”, IEEE 19th India Council International Conference (INDICON), Kochi, India, 2023, doi: https://doi.org/10.1109/INDICON56171.2022.10039789.

[17] S. Nikalje and M. V. Mane, “Copy-Move and Image Splicing Forgery Detection based on Convolution Neural Network”, International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT), Kannur, India, 2022, doi: https://doi.org/ 10.1109/ICICICT54557.2022.9917679.

[18] D. Mallick, M. Shaikh, A. Gulhane and T. Maktum, Copy Move and Splicing Image Forgery Detection using CNN, ITM Web Conferences, vol. 44, pp. 03052-03058, 2022, doi: https://doi.org/10.1051/itmconf/20224403052.

[19] C. Dai, L. Su, B. Wu and J. Chen, DS-Net: Dual supervision neural network for image manipulation localization, IET Image Processing, 2023, doi: https://doi.org/10.1049/ipr2.12885.

[20] R. Xiang, G. Liu, K. Li, J. Liu, Z. Zhang and M. Dang,” Zero-watermark scheme for medical image protection based on style feature and ResNet”, Biomedical Signal Processing and Control, vol. 86, nº A, pp. 105127, 2023, doi: https://doi.org/10.1016/j.bspc.2023.105127.

[21] F. Dong, J. Li, U. A. Bhatti, J. Liu, Y.-W. Chen and D. Li, “Robust Zero Watermarking Algorithm for Medical Images Based on Improved NasNet-Mobile and DCT”, Electronics, vol. 16, nº 16, pp. 3444, 2023, doi: https://doi.org/10.3390/electronics12163444.

[22] C. Li, H. Sun, C. Wang, S. Chen, X. Liu, Y. Zhang, N. Ren and D. Tong, "ZWNet: A Deep-Learning-Powered Zero-Watermarking Scheme", Applied Sciences, vol. 14, pp. 435, 2024.

[23] L. He, Z. He, T. Luo and Y. Song, "Shrinkage and Redundant Feature Elimination Network-Based", Symmetry, vol. 15, nº 5, pp. 964, 2023, doi: https://doi.org/10.3390/sym15050964.

[24] Q. Dong, L. Feng and T. Lu, "Reversible Watermarking Algorithm for Halftone Images Based on Overlapping Blocks Scanning and Central Pixels Flipping", International Conference on Computer Graphics, Image and Virtualization (ICCGIV), Chongqing, China, 2023, doi: https://doi.org/10.1109/ICCGIV57403.2022.00054.

[25] W. Lu, Y. Xue, Y. Yeung, H. Liu, J. Huang and Y.-Q. Shi, "Secure Halftone Image Steganography Based on Pixel Density Transition", IEEE Transactions on Dependable and Secure Computing, vol. 18, nº 3, pp. 1137 - 1149, 2021, doi: https://doi.org/10.1109/TDSC.2019.2933621.

[26] I. Amerini, L. Ballan, R. Caldelli, A. D. Bimbo and G. Serra, A SIFT-based forensic method for copy-move attack detection and transformation recovery, IEEE Transactions on Information Forensics and Security, vol. 6, nº 3, pp. 1099-1110, 2011, doi: https://doi.org/ 10.1109/TIFS.2011.2129512.

[27] J. Dong, W. Wang and T. Tan. "CASIA Image Tampering Detection Evaluation Database", IEEE China Summit and International Conference on Signal and Information Processing, Beijing,China, 2013, doi: https://doi.org/10.1109/ChinaSIP.2013.6625374.

[28] N. T. Pham, J.-W. Lee, G.-R. Kwon and C.-S. Park, "Hybrid Retrieval Method for Image Splicing Validation", Symmetry, vol. 11, n° 1, pp. 83, 2019, doi: https://doi.org/10.3390/sym11010083.

Eating Behavior and Level of Knowledge About Healthy Eating Among Gym Users: A Multinomial Logistic Regression Study

Ana Huamani-Huaracca[1], Sebastián Ramos-Cosi[2], Michael Cieza-Terrones[3],
Gina León-Untiveros[4], Alicia Alva Mantari[5*]
Department of E-Health Research, Faculty of Health Sciences, Universidad de Ciencias y Humanidades, Lima, Peru[1]
Image Processing Research Laboratory (INTI-Lab), Universidad de Ciencias y Humanidades, Lima Peru[2, 5]
Facultad de Medicina Alberto Hurtado, Universidad Peruana Cayetano Heredia, Lima, Peru[3]
Universidad Privada Norbert Wiener, Lima, Peru[4]

***Abstract*—The World Health Organization indicates that unhealthy diets cause approximately 11 million deaths annually worldwide. In Peru, 57.9% of the population consumes highly processed foods daily. The objective of this study is to analyze the relationship between knowledge about healthy eating and eating behavior among gym users in a district of Lima, Peru. Using an exploratory and quantitative design, information was collected from 156 users through a hybrid questionnaire, analyzed with SPSS and multinomial logistic regression techniques. The results reveal that 57.42% of the participants have an intermediate knowledge of healthy eating, while only 17.42% reach a high level. Likewise, 49.03% exhibit an intermediate eating behavior. In addition, sociodemographic factors, such as the duration of gym attendance and maintenance of a specific diet, were found to influence eating behavior. It is concluded that there is a significant relationship between the level of knowledge and eating behavior, underlining the importance of nutrition education to improve eating habits in this population.**

***Keywords*—Knowledge; healthy eating; gym; eating behavior**

I. INTRODUCTION

The World Health Organization (WHO) points out that, in the world, the maintenance of unhealthy diets causes about 11 million deaths each year [1] and eight million deaths are registered as a result of an unhealthy diet [2]. In addition, it is identified that more than 40% of the world's population has inadequate knowledge about healthy eating [3]. In countries such as the United States, Saudi Arabia and Spain, 87.72% of people do not practice healthy eating habits, a situation that represents a global problem [4].

In South America, Peru is the most food-insecure country with 16.6 million people exposed [5]. The National Institute of Health (INS) warns that 57.9% of Peruvians consume foods with a high degree of processing and added chemicals on a daily basis [6]. In this context, food consumption is influenced by society and by the lack of knowledge of food practices. In Huancayo, only 17.4% of gym users know how to interpret food nutrition labels [7].

During the last few years, people have developed a need to balance physical and psychological health, evidence of this is the insertion in sports activities [8]. In Peru, an average of 5.2 hours of exercise per week is practiced and 28% of these activities correspond to gyms [9]. However, there are barriers that prevent sports practice such as lack of time in almost half of the population [10]. On the other hand, 44% of Lima residents do physical activity frequently [11].

Healthy eating is based on the consumption of healthy, adequate and pleasant foods, to meet nutritional needs and preserve health [12]. In addition, it involves knowing the composition of food and executing practices that ensure adequate nutrition [13]. In Lima, inconsistencies have been detected in the perception of healthy eating, 1 in 2 residents believe that it is based on consuming vegetables and 39% consider that it is avoiding the consumption of harmful foods [11].

There are several reasons that prevent the implementation of healthy eating, in gym users there is a lack of knowledge to analyze the products before consuming them, 27.5% of them frequently read the nutritional information and focus mainly on the level of fat and proteins, this is how less interest is given to the nutrients of the food [7]. In Peruvian society, 1 in 2 people lives at food risk, detecting poverty as a mitigating factor represented with an impact of 25.9% at the national level [5].

Inequalities in the knowledge and practice of healthy eating behaviors represent a latent risk, which shows that in Peruvian households 33% of food expenditures are made outside the home and on unhealthy foods [14]. A much bigger problem is the development of eating disorders, which is frequently identified in people who exercise with the goal of losing weight [15]. In addition, poor diet favors the development of non-communicable diseases such as diabetes and the susceptibility of acquiring a communicable disease due to the weakening of the immune system [16].

Therefore, the main objective of this study is to analyze the influence of the level of knowledge about healthy eating on eating behavior, through the collection of information from gym users in a district of Lima, to contribute to the preservation of their health. In this way, it is particularly intended to determine the level of knowledge about healthy eating and eating behaviors through a hybrid questionnaire that integrates the Healthy Eating Knowledge Test and the Eating Behavior

Questionnaire, to then correlate the level of knowledge and healthy eating behaviors through a comparative analysis. In addition, it seeks to provide information to promote knowledge and improve behaviors in gym users.

This study is organized into sections that facilitate information management. First, a literature review is conducted, describing previous works of similar scope to contextualize the topic in Section II. Then, the description and formulation of the problem are presented, highlighting its relevance in Section III. Next, the materials and methods are presented, detailing the resources and procedures used in Section IV. Section V highlights the key findings for the subsequent stages. In the discussion, the results are interpreted and compared with previous studies, allowing for the identification of relevant similarities or differences in Section VI. Finally, Section VII offers a reflection on the entire study.

II. Literature Review

In the study by Başpinar et al. [17], the influence of nutritional knowledge and practices on the quality of nutrition in people who maintain regular physical activity was addressed in order to determine the association between diet quality, nutritional knowledge and eating practices. The research was quantitative because they applied the Healthy Eating Index-2015 questionnaire and qualitative because they added interviews with 200 users of regular gyms. The results obtained show that 47.0 % of the sample had poor nutritional knowledge and 48.0 % inadequate nutritional practices; In addition, nutritional knowledge is not related to diet quality or food consumption in general ($p<0.05$).

The literature review conducted by Maza et al. [4] on eating habits in university students using the Systematic Review of the Literature and following the PRISMA standards (Preferred Reporting Elements for Systematic Reviews and Meta-Analyses), was composed of 57 articles. A prevalence of inappropriate dietary practices was found and they are characterized by incorrect information management; they are reflected in the consumption of unhealthy foods, easily available and with a minimum amount of nutritional use. Likewise, unhealthy behaviors were detected in the diet of young people; with this comes the risk of suffering health problems in the short term such as being overweight or underweight, and in the long term much greater complications such as cardiovascular diseases and diabetes.

During Reyes and Oyola's research [18], the level of knowledge about healthy eating was determined in university students of various professional careers, through a questionnaire developed with the Kuder-Richardson reliability measure and applied to 136 people. With the results, it was stated that 41.2% of university students have an average food knowledge and students of health careers stand out with a higher level. Therefore, the advantage of knowing more widely how to guide eating behaviors suggests greater efficiency in preserving their health. In this sense, it is pointed out that health promotion activities should focus on providing strategies to positively influence university students.

The study by Damián et al. [7] on the knowledge of gym users in the management of nutritional labels aimed to demonstrate that using and interpreting food labels influences nutrition. The instrument used was a test composed of the variables reading, use and interpretation; which was aimed at 385 users. The information obtained shows that the participants had an average age of 27.8 ± 9.3 years, 44.7% were men and that only 17.4% correctly handle the nutritional information on the labels, so it is important to promote education in this population.

With Wanden's research [19] on the role of diet in the proper functioning of the body and the response to diseases, a study on nutrition was undertaken, prioritizing older people because they are susceptible to instability between the consumption and energy expenditure of nutrients. The proposed approach was nutritional screening, which was based on the identification of people at risk of malnutrition with the aim of applying a nutritional assessment to this group in which potentially harmful alterations are identified early. Among the most important indicators were weight, height and functional assessment; these in turn made it possible to recognize vulnerable people who received timely treatment.

In the project of Fuentes et al. [20] the objective was to demonstrate that there are dietary factors that participate in the growth and development of sports gymnasts, so a case study was carried out in phases in an adolescent athlete with signs of inadequate nutrition and stunted growth. First, they were provided with education, then balanced foods were inserted into their diet, preventive macro and micronutrients were included, and finally the amount of nutritional supplements was adjusted according to the evolution of their needs. The results collected indicated that the adolescent presented improvements in his growth, development and physical performance. Stress and excessive sports demands were identified as the main causes of eating problems, alterations in body image and different health problems.

Researchers Marquez et al. [21] specified in a study on eating behavior in university students that this group is more likely to develop alterations due to the lack of information and willingness to preserve their health. To measure eating behaviors, they designed a questionnaire composed of 31 questions, which after being validated was applied to 333 students. The results indicated that most students have a medium tendency to have healthy eating habits and evaluating eating behavior made it possible to plan interventions. The designed instrument demonstrated high effectiveness and validity to be applied in studies that aim to analyze eating behavior.

III. Description and Formulation of the Problem

Healthy eating and its adherence to daily life represents a public health problem. Eating behaviors directly influence the state of health of the human being, it has been shown that healthy eating is beneficial in the medium and long term because it improves the state of health and prevents some diseases [2]. Food goes through a series of stages before reaching our tables in conventional presentation, the selection of quality foods for nutrition intervenes in the preservation of health [22]. However, despite the fact that proper nutrition is a

right and a basic need, unhealthy eating is prevalent in modern society [23].

Knowledge about healthy eating is of fundamental importance, especially for the population that maintains a healthy lifestyle. In this sense, people who attend gyms constitute a key group as they are influenced by the competitiveness of their discipline and different dietary factors [7]. It is identified that among the main motivations of gym users to go to fitness centers are the maintenance of health, body image and competition [24]. On the other hand, eating disorders are even more prevalent in people who practice endurance sports [25].

The relevance of our study in gym users is established because they are an underserved group and will be theoretically provided by determining the relationship between the level of knowledge and healthy eating. In addition, at the methodological level, the validation of the integrated instrument and the development of a plan structured by stages that facilitate the process will be contemplated. Likewise, in practice, relevant information will be collected to process it and transform it into percentages that indicate trends in relation to the indicators. In the social sphere, the availability of information will be facilitated to promote knowledge, reduce risk behaviours and contribute to the 2030 Agenda with specific support for the Sustainable Development Goal "Good health and well-being" [26].

IV. Materials and Methods

The materials and methods should be described in sufficient detail to allow the reader to understand the methodological process that has been followed. To this end, the following subsections must be completed:

A. Design and Approach

The study will have an exploratory design because the objective is focused on evaluating a problem that has been little studied in the population made up of gym users and will bring us closer to obtaining new knowledge [27]. In addition, it will have a quantitative approach that will allow the controlled measurement of data for the processing and generalization of trends [28].

B. Population, Sample and Sampling

Due to the massive increase in gyms and fitness centers in Lima, the district of Comas is selected to make use of the services of Gymfinder Group and the presence of approximately 32 gyms in the district is determined [29]. Then, from the Stylos Fitnes and Urban Gym centers, it is estimated that, per gym, there are approximately 80 people enrolled. This amount is multiplied by the number of gyms in the district and the result obtained is a population of 2560 users.

1) Sample size: The determination of the sample was processed by means of the statistical program "EPIDAT 4.2" in which a population of 2560 users was included, with an expected proportion of 87.72% [4], a confidence level of 95% and a design effect of 1; resulting in a sample of 156 users.

[1] Tamaño de muestra. Proporción:

Datos:

Tamaño de la población:	2,560
Proporción esperada:	87,720%
Nivel de confianza:	95,0%
Efecto de diseño:	1,0

Resultados:

Precisión (%)	Tamaño de la muestra
5,000	156
10,000	41

Fig. 1. Epidat 4.2 sample calculation.

Fig. 1 shows the result of the sample based on the process carried out by the statistical program Epidat 4.2 and the result obtained is 156 users.

2) Sample selection: A simple random sampling was carried out, including criteria such as the location of the establishments, ease of access and availability for intervention.

3) Inclusion criteria:

- Users of gyms in the district of Comas.
- Users enrolled in a gym in the district of Comas.
- Users over 18 years of age.
- Users of both sexes.
- Users who accept informed consent

4) Exclusion criteria:

- Users of gyms in districts not located in Comas.
- People who are not registered in a gym in the district of Comas.
- Users under 18 years of age.
- Users who do not accept informed consent.

C. Study Variables

The variables considered in the research are chosen because they are related to each other for the comprehensive assessment of the participants and make it possible to identify trends regarding possible alterations.

1) Knowledge about healthy eating:

a) Conceptual definition: Knowledge about healthy eating is the amount of information that each person has to justify their dietary behaviors and is of vital importance for the preservation of health. Its availability determines the behaviors that are developed in the diet, established foods, etc. [18].

b) Operational definition: Knowledge about healthy eating is the information that is acquired in society, its development and organization influences the eating habits of

gym users in Lima. In this way, knowledge about healthy eating intervenes in the preservation of health.

2) Eating behavior:

a) Conceptual definition: Eating behavior is the relationship that is established between human beings and food, it is a set of actions that influence daily food. It is directly involved in food consumption, nutrient absorption and health status [21].

b) Operational definition: Eating behavior is the way in which eating behaviors are oriented in daily life, it is defined by eating habits, handling, selection and consumption of food. Its alteration determines the nutritional condition and health status of gym users in Lima.

D. Measuring Technique and Instrument

1) Data collection technique: The data collection process was carried out through the application of an integrated questionnaire with sections extracted from two validated tools, each of them focused on the individual study variables, so its constitution is of great relevance.

a) For the variable knowledge about healthy eating: The questionnaire called "Knowledge test on healthy eating" was selected, which is used to measure the level of knowledge and was developed by Reyes and Oyola [18] following the parameters established by the WHO and other authors. It is made up of 2 dimensions, the first collects general data on age and sex and the second corresponds to knowledge about healthy eating with 16 questions. For the evaluation, a value of 2 is scored for the correct answers and a total classification of: <17 "low", 17-25 "medium" and > 25 "high".

- *Validity and reliability:* The reliability of the instrument was determined based on the Kuder Richardson coefficient with a reliability value of 80.7% classified as acceptable and its validation is proven in the Peruvian population.

b) For the variable eating behavior: The "Eating Behavior Questionnaire" was chosen, consisting of 31 questions distributed in eight dimensions and was designed by Márquez et al. [21]. The dimensions that compose it are meal times, food and beverage consumption preferences, way of preparing food, reading nutritional labels, consumption of food outside the home, satiety, following therapeutic or special diets, and perception of healthy eating, barriers to change and beliefs, each item is evaluated from 1 to 5 according to the Likert scale.

- *Validity and reliability:* The level of reliability was calculated according to the evaluation of 15 health experts in the clinical and research area. The validity of the tool was established abroad, due to which its reproducibility in our country and in the target population was validated by trained professionals from the local environment.

2) Hybrid sizing: Based on the aforementioned tools and the dimensions that correspond to each of them, in order to satisfy all the objectives set and optimize the collection of information, a hybrid dimensioning was carried out by summarizing the aspects that most cohesive the information obtained during the collection of relevant data for our study. Therefore, it was decided to use the Healthy Eating Knowledge Test because it is the one that offers a greater approximation to the main variable of the study, excluding questions 10 and 16 from the instrument because they have shown ambiguity in the results of their respective study. In addition, the first section of the Eating Behavior Questionnaire is taken exclusively composed of 8 items, as it is the one that offers the greatest opportunity to collect data that complement the defined objectives, excluding questions 5 and 7. Finally, after a detailed analysis of the effectiveness of the indicators of each of the instruments in their respective backgrounds, a total of 20 items are established for the instrument of this study. For the evaluation of this new instrument, a score of 1 for each of the correct answers and a total classification of: <11 "low", 11-15 "medium" and > 16 "high" are considered.

TABLE I. HYBRID INSTRUMENT SIZING

Healthy eating knowledge test	Knowledge about healthy eating	14 items
Eating Behavior Questionnaire	Eating behaviors	6 items

Table I shows the hybrid instrument designed to efficiently integrate the practical dimensions of each of the tools studied, this construction was evaluated by five national experts in health and scientific research, an evaluation with which a final Anyken V of 0.93 was obtained, which qualifies the integrated tool as valid for its application. On the other hand, to determine the reliability of the instrument, a pilot test was carried out with 20 users; then, the information obtained was processed using the Kuder Richardson coefficient – KR20 as it is the ideal one to analyze dichotomous variables, with which a total value of 0.86 was found, cataloguing the instrument as good in terms of reliability.

E. Procedure for Data Collection

1) Bio-ethical aspects: This project prioritizes respect for the bioethical principles of autonomy, beneficence, non-maleficence and justice; They contain aspects that ensure the protection of the participant [30]. Prior to the execution of the visualized fieldwork, the approval of the ethics committee is required, so the product was directed to the responsible committee for evaluation by experts in the field of research and approval was obtained for the application of the study in the selected field.

- *Autonomy:* By distributing informed consent as the first step of intervention so that users can freely choose to participate in the study.
- *Beneficence:* It will be used to provide information that contributes to the strengthening of knowledge about healthy eating in gym users.
- *Non-maleficence:* By establishing a bond with each person to facilitate explanation and understanding

regarding the absence of questions of a personal nature that may transgress the individuality of the participants.

- *Justice:* Evidenced by the egalitarian approach of all members without differences of any kind.

2) Informed consent: Informed consent represents research ethics, ensuring respect for the principle of autonomy of all those who, by their own decision, decide to participate in the study. With this document, participants will have the assurance that their rights and privacy will not be violated and will be used for strictly academic purposes. During the fieldwork, it will be provided physically before starting with the application of the designed instrument.

3) Instrument application: During the first two weeks of May 2024, the entry to the physical conditioning establishments that belong to the district of Comas was coordinated through an official letter addressed to the management demonstrating the approval by the ethics committee and the material to be evaluated according to the inclusion and exclusion criteria. Then, the instrument was applied on the premises allowing the development of the study.

The acquisition of information was directly through a virtual form distributed through a link sent to the telephone application of convenience per user at the time of the intervention. If an electronic device or an appropriate application was not available at the time of the study, participation was carried out from the interviewers' mobile phones.

4) Data analysis: For the storage of the information extracted during the research, the Google Sheets program was used for its easy handling and the organization it has to quantify what was obtained [31]. On the other hand, the data was processed in the SPSS software that specializes in advanced statistical analysis determining the correlation of Spearman's coefficients between the level of knowledge about healthy eating and eating habits [32]. Likewise, to establish the reliability of the instrument, Cronbach's alpha coefficient was calculated based on the questionnaires already validated by this same statistical technique.

F. Data Processing

1) Multinomial logistic regression model: This tool is used to predict the outcome of a variable categorized as dependent with a large number of categories. This model is characterized by reducing the extension limitations found in other tools, since it allows the dependent variable to take two or more categories. Their inclusion in this study is important for the management of categorical variables at multiple levels because it allows us to understand in detail the influence of predictor variables on the probability of belonging by category [33].

V. RESULTS

Table II shows that 62.18% (97) of the participants were between 18 and 27 years old, 19.87% (31) were between 28 and 37 years old, 14.74% (23) were between 38 and 47 years old, and only 0.64% (1) were between 58 years and older. With respect to sex, 53.85% (84) are women and 46.15% (72) are men. Regarding the time spent attending the gym, 26.28% (41) attended 1 month or less ago, 27.56% (43) from 2 to 6 months, 3.21% (5) from 7 to 11 months, 17.95% (28) for 1 year or more and 25% (39) attended non-continuously. Regarding the time elapsed since attending a nutritionist, 20.51% (32) have had guidance 1 month or less ago, 18.59% (29) have had guidance from 2 to 6 months ago, 7.69% (12) from 7 to 11 months, 6.41% (10) have been assisted by a nutritionist 1 year or more ago, and 46.8% (73) have not been assisted by a nutritionist. In relation to the maintenance of a specific diet, 25% (39) have maintained it for 1 month or less, 32.5% (50) for 2 to 6 months, 12.82% (20) for 1 year or more and 30.13% (47) do not maintain a specific diet. The origin of nutritional information in 58.97% (92) is from social networks, 25.64% (40) obtain it from a nutritionist, 8.33% (13) from friends and 7.06% (11) from family members.

TABLE II. SOCIODEMOGRAPHIC DATA OF GYM USERS OF A COMAS GYM

Sociodemographic data	n= 156	
	fi	%
Age		
18 - 27	97	62,18
28 - 37	31	19,87
38 - 47	23	14,74
48 - 57	4	2,57
58 years or older	1	0,64
Sex		
Female	84	53,85
Male	72	46,15
Gym attendance		
1 month or less	41	26,28
2 to 6 months	43	27,56
7 to 11 months	5	3,21
1 year or more	28	17,95
Non-Continuous Support	39	25,0
Assistance to nutritionist		
1 month or less	32	20,51
2 to 6 months	29	18,59
7 to 11 months	12	7,69
1 year or more	10	6,41
There was no attendance	73	46,80
Maintaining a specific diet		
1 month or less	39	25,0
2 to 6 months	50	32,05
7 to 11 months	0	0,0
1 year or more	20	12,82
Doesn't maintain a diet	47	30,13
Source of Nutrition Information		

Social Media	92	58,97
Nutritionist	40	25,64
Friends	13	8,33
Family	11	7,06
University	0	0,0

A. Sample Setting

During the organization of the data, it has been possible to detect that one of the 156 intervened users has provided insufficient information to be analyzed in the study results, a situation for which it has been decided to exclude him from the next stage, therefore, the sample processed later will be made up of a total of 155 participants.

On the other hand, for the analysis of the data by age, the current classification of age groups according to the WHO has been applied, which has allowed the distribution of the sample data in a group composed of young people aged 18 to 26 years and another made up of adults between 27 and 59 years of age [34], [35].

B. Level of Knowledge and Behavior by Gender

TABLE III. LEVEL OF KNOWLEDGE ABOUT HEALTHY EATING BY GENDER

Level of knowledge	Value	Gender		Total
		Female	Male	
Low	n	20	19	39
	%	12,9	12,26	25,16
Intermediate	n	51	38	89
	%	32,9	24,52	57,42
High	n	12	15	27
	%	7,74	9,68	17,42
Total	n	83	72	155
	%	52,55	46,45	100

Table III shows that 12.9% (20) women and 12.26% (19) men have a low level of knowledge about healthy eating, representing 25.16% (39) of the total. The intermediate level of knowledge is present in 32.9% (51) women and 24.52% (38) men, which constitute 57.42% (89) of the total. Finally, 7.74% (12) females and 9.68% (15) males make up the high level of knowledge, an amount that represents 17.42% (27) of the total.

TABLE IV. EATING BEHAVIOR BY GENDER

Eating behavior	Value	Gender		Total
		Female	Male	
Low	n	23	28	51
	%	14,84	18,06	32,9
Intermediate	n	47	29	76
	%	30,32	18,71	49,03
High	n	13	15	28
	%	8,39	9,68	18,06
Total	n	83	72	155
	%	53,55	46,45	100

Table IV presents healthy eating behavior by gender, where low eating behavior is manifested in 14.84% (23) women and 18.06% (28) men, which corresponds to 32.9% (51) of the total. The intermediate behavior is manifested by 30.32% (47) women and 18.71 (29) men, which represents 49.03% (76) of the total. Finally, 8.39% (13) females and 9.68% (15) males make up 18.06% (28) of the total.

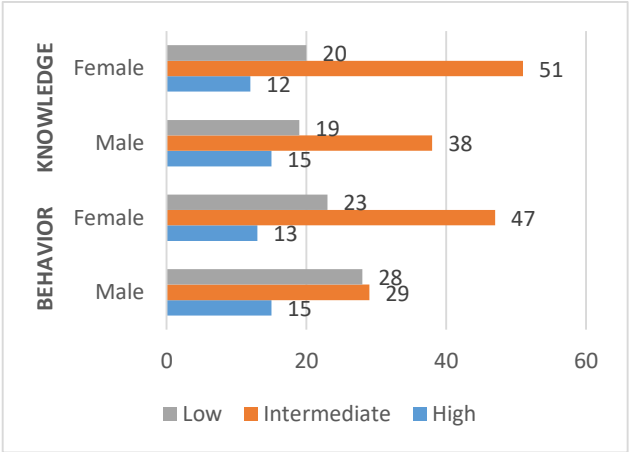


Fig. 2. Relationship between the level of knowledge and behavior according to gender.

Fig. 2 shows that high male eating behavior is present in only 15 users, intermediate in 29 and low in 28. In the female sex, there was a high level of behavior in 13 people, intermediate in 47 and low in 23. Likewise, the male knowledge about healthy eating is made up of 15 users with high knowledge, 38 with an intermediate level and 19 with low knowledge. In the female sex, high knowledge was detected in 12 participants, intermediate knowledge in 51 and a low level in 20. These results demonstrate the predominance of the intermediate level of knowledge and behavior over the low and high.

C. Level of Knowledge and Behavior by Age

TABLE V. LEVEL OF KNOWLEDGE ABOUT HEALTHY EATING BY AGE

Level of knowledge	Value	Age		Total
		Young	Adult	
Low	n	25	14	39
	%	16,13	9,03	25,16
Intermediate	n	59	30	89
	%	38,06	19,35	57,42
High	n	13	14	27
	%	8,39	9,03	17,42
Total	n	97	58	155
	%	62,58	37,42	100

Table V shows that 16.13% (25) young people and 9.03% (14) adults have a low level of knowledge about healthy eating, representing 25.16% (39) of the total. The intermediate level of knowledge is present in 38.06% (59) young people and 19.35% (30) adults, which constitute 57.42% (89) of the total. Finally,

8.39% (13) young people and 9.03% (14) adults make up the high level of knowledge, an amount that represents 17.42% (27) of the total.

TABLE VI. EATING BEHAVIOR BY AGE

Eating behaviour	Value	Age		Total
		Young	Adult	
Low	n	41	10	51
	%	26,45	6,45	32,9
Intermediate	n	35	41	76
	%	22,58	26,45	49,03
High	n	21	7	28
	%	13,55	4,52	18,06
Total	n	97	58	155
	%	62,58	37,42	100

Table VI shows eating behavior by age, where low eating behavior is manifested in 26.45% (41) young people and 6.45% (10) adults, which corresponds to 32.9% (51) of the total. The intermediate behavior is represented by 22.58% (35) young people and 26.45% (41) adults, which makes up 49.03% (76) of the total. Finally, 13.55% (21) young people and 4.52% (7) adults make up 18.06% (28) of the total.

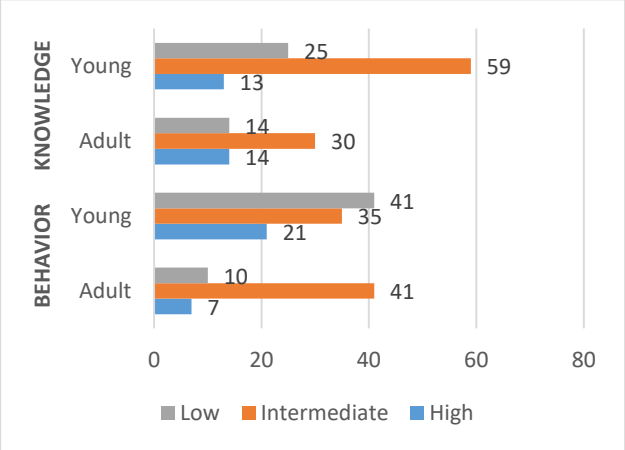


Fig. 3. Relationship between the level of knowledge and behavior according to age.

In Fig. 3 it can be seen that adult eating behavior is high in only seven participants, intermediate in 41 and low in 10. The young population is made up of 21 people with high behavior, 35 with intermediate behavior and 41 with low behavior. In addition, knowledge about healthy eating in adults is represented by 14 users with a high level, intermediate knowledge with 30 and low knowledge with 14. In young people, the level of knowledge is high in 13 people, intermediate in 59 and low in 25. These data affirm the trend of the intermediate level of knowledge and behavior over the low and high.

D. Application of the Multinomial Logistic Regression Model

The model has been applied to analyse the factors that influence eating behaviour, specifically considering the level of knowledge and the most relevant socio-demographic data, the process has been defined using the following formula:

$$C = 0.1891033 + 0.8314438(V1) - 0.0868413(V2) + 0.1306843(V3) - 0.0927213(V4)$$

Where:

C: Eating behavior

V1: Knowledge Level

V2: Time of attendance with a nutritionist

V3: Time in which you maintain a specific diet

V4: Continuous Gym Attendance Time

By means of this calculation, it can be stated that the Multinomial Logistic Regression Model is acceptable, since it has a p value of 0.00 demonstrating its significance and an adjusted Spearman's Rho correlation coefficient level of 0.54.

TABLE VII. PROBABILITY OF CONTRAST OF THE LOGISTIC MODEL FOR BEHAVIOR BASED ON THE LEVEL OF KNOWLEDGE

p value (Sig.)		Eating behaviour	
		Intermediate	High
Level of knowledge	Intermediate	0,99	0,994
	High	0,0	0,996

p value = probability that is used to contrast with the level of significance or alpha value (α=0.05).

Table VII presents the association between the level of knowledge about healthy eating and the eating behavior of the participants. It can be detected that the p value (Sig.) is greater than 0.05 in the variables, except for the relationship between high level of knowledge and intermediate behavior; which statistically shows that the greater the knowledge, the greater the risk of having an intermediate behavior.

TABLE VIII. PROBABILITY OF CONTRAST OF THE LOGISTIC MODEL FOR BEHAVIOR BASED ON SOCIODEMOGRAPHIC DATA

p value (Sig.)		Eating behaviour	
		Intermediate	High
Sociodemographic data	Nutritionist Assistance 1 month or less	0,462	0,328
	Specific diet for 2 to 6 months	0,01	0,994
	Gym attendance 1 month or less ago	0,00	0,00

p value = probability that is used to contrast with the level of significance or alpha value (α=0.05).

Table VIII shows the relationship between the most significant sociodemographic data and the eating behaviour of users. We observed that the p value (Sig.) is less than 0.05 in three of the established associations. First of all, statistics show that maintaining a specific diet for two to six months means a higher risk of having an intermediate behavior. Continuous attendance for 1 month or less is associated with a lower risk of intermediate behavior. Finally, attending the gym for 1 month

or less implies a lower risk of maintaining a high eating behavior.

VI. Discussion

It has been calculated that the level of knowledge about healthy eating is low in 25.16% of gym users in Lima and inappropriate eating behavior is present in 32.9%. However, in Turkey, Başpinar et al. [17] report that 47.0% of their sample had poor nutritional knowledge and 48.0% inadequate nutritional practices. This differentiation could be caused by the geographical location that is so different in both studies and the sociodemographic mitigating factors of each country. In addition, the review by Maza et al. [4] has detected the trend of unhealthy behaviors in young people, in contrast to our study, in which 57.42% tend to present intermediate eating behaviors.

In this research, it is found that 57.42% of gym users have an intermediate knowledge about healthy eating, an amount higher than the 41.2% average knowledge established by Reyes and Oyola [18] in their research on university students. This distinction may be due to the fact that in gyms there are a greater number of people looking for information to maintain a good physical condition, while in universities young people focus more on academic knowledge. In relation to eating behaviour, it has been shown that 49.03% of gym users have an intermediate behaviour, similar to the study by Marquez et al. [21], in which a medium trend is also detected.

In our study, it has been found that 62.58% of the users are between 18 and 26 years old, males make up 46.45% and 17.42% have a high knowledge of healthy eating; data that agree with the research of Damián et al. [7], in which a mean age of 27 years has been obtained, male participants represented 44.7% and only 17.4% correctly handled the information on nutritional labels. This similarity shows that in cities such as Lima and Huancayo (coast-highlands) the young population tends to attend fitness centers, women with a slight frequency and that knowledge about healthy eating is high in only a small percentage of users, which makes them susceptible to inappropriate eating behaviors.

With respect to the Multinomial Logistic Regression Model, it has been possible to identify a p value of 0.00 and an adjusted Spearman's Rho of 0.54, which has shown that eating behavior is influenced by the level of knowledge about healthy eating, the time spent attending a nutritionist, the time spent on a specific diet and the time spent continuously attending the gym. This transcendental finding differs from what was found in similar studies such as [17] in which the relationship between nutritional knowledge and the consumption of foods with a p value less than 0.05 is denied.

On the other hand, during the execution of this study, there have been limitations related to the intervention schedule of the participants, since a low average influx of users available for the intervention was evidenced, which is why the moment of greatest agglomeration for each gym had to be observed for several days. Likewise, some of the users gave up participating due to lack of time, being in the middle of training or other activities of similar scope. Finally, some gyms prevented the interviewer from passing through due to internal policies and other centers had to be resorted to.

VII. Conclusion

In conclusion, the present research determined a p value of 0.00 and an adjusted Spearman's Rho of 0.54 using the Multinomial Logistic Regression Model, which provides sufficient statistical evidence to affirm that there is a direct relationship between the level of knowledge about healthy eating and eating behavior. Specifically, people with high knowledge are more likely to maintain intermediate behavior. Among gym users in Lima, an intermediate level of knowledge and behavior predominates with 57.42% and 49.03% respectively.

In addition, sociodemographic factors have shown an association between the maintenance of a specific diet for 2 to 6 months and intermediate eating behavior. Attending the gym 1 month or less ago indicates a lower risk of maintaining intermediate or high behavior. Finally, it is observed that female users and young people tend to have greater knowledge about healthy eating and better eating behaviors.

It is highlighted that the design of the hybrid questionnaire to measure knowledge about healthy eating and eating behaviors was fundamental to analyze the situation in which gym users in Lima find themselves. The validity and reliability of this instrument have been evaluated by national experts and its usefulness for future studies of similar scope has been demonstrated.

Finally, the information obtained in the present study is available for future research, providing a solid basis for the generation of new knowledge and the validation of results. This information can serve as a reference for comparisons in different regions of the country and the world, allowing the consistency of the findings to be evaluated in different geographical and cultural contexts.

References

[1] "Summary version of The State of Food Security and Nutrition in the World 2022," Summary version of The State of Food Security and Nutrition in the World 2022, Jul. 2022, doi: 10.4060/CC0640ES.

[2] M. T. Escalé et al., "[Nutritional education in school-age children through Programa Nutriplato®]," Nutr Hosp, vol. 37, no. Spec No2, pp. 47-51, 2021, doi: 10.20960/NH.03357.

[3] "UN report: The number of hungry people in the world increased to 828 million in 2021 - PAHO/WHO | Pan American Health Organization." Accessed: Jun. 21, 2024. [Online]. Available: https://www.paho.org/es/noticias/6-7-2022-informe-onu-numero-personas-que-padecen-hambre-mundo-aumento-828-millones-2021

[4] F. J. Maza-Ávila, M. C. Caneda-Bermejo, A. C. Vivas-Castillo, F. J. Maza-Ávila, M. C. Caneda-Bermejo, and A. C. Vivas-Castillo, "Eating habits and their effects on the health of university students. A systematic review of the literature," Psychogens, vol. 25, no. 47, pp. 110–140, Jan. 2022, doi: 10.17081/PSICO.25.47.4861.

[5] "Food and Agriculture Organization of the United Nations: Peru is the country with the highest Food Insecurity in South America | FAO in Peru | Food and Agriculture Organization of the United Nations." Accessed: Jun. 21, 2024. [Online]. Available: https://www.fao.org/peru/noticias/detail-events/es/c/1603081/

[6] "The largest amount of salt consumed by Peruvians comes from processed foods - News - Ministry of Health - Peruvian State Platform." Accessed: Jun. 21, 2024. [Online]. Available: https://www.gob.pe/institucion/minsa/noticias/46032-la-mayor-cantidad-de-sal-que-consumen-los-peruanos-proviene-de-alimentos-procesados

[7] N. Damián-Bastidas, R. J. Chala-Florencio, R. Chávez-Blanco, and P. Mayta-Tristán, "Reading, Use, and Interpretation of Nutrition Labels in Huancayo Gym Users, Peru 2015," Nutr HospVol. 33 No. 6 pp. 1410–1417, 2016, doi: 10.20960/n.803.

[8] N. I. Vargas-Cuentas, S. J. Ramos-Cosi, and A. Roman-Gonzalez, "Air Quality Analysis in Pasco Peru Using Remote Sensing," International Journal of Engineering Trends and Technology, vol. 69, no. 12, pp. 30–38, Dec. 2021, doi: 10.14445/22315381/IJETT-V69I12P205.

[9] "National Survey 2021 - Campaigns - Peruvian Sports Institute - Peruvian State Platform." Accessed: Jun. 21, 2024. [Online]. Available: https://www.gob.pe/institucion/ipd/campa%C3%B1as/11524-encuesta-nacional-2021

[10] "Global views on sports: 58% globally would like to practice more | Ipsos." Accessed: Jun. 21, 2024. [Online]. Available: https://www.ipsos.com/es-mx/puntos-de-vista-globales-sobre-los-deportes-al-58-nivel-mundial-le-gustaria-practicar-mas

[11] "Food and healthy living in Lima | Ipsos." Accessed: Jun. 21, 2024. [Online]. Available: https://www.ipsos.com/es-pe/alimentacion-y-vida-saludable-en-lima

[12] E. Waters et al., "Interventions for preventing obesity in children," Cochrane Database of Systematic Reviews, vol. 2011, no. 12, p. CD001871, Dec. 2011, doi: 10.1002/14651858.CD001871.pub3.

[13] S. Yorde Erem, "How to Achieve a Healthy Life," Venezuelan Annals of Nutrition, vol. 27, no. 1, pp. 129–142, 2014, Accessed: Jun. 21, 2024. [Online]. Available: http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0798-07522014000100018&lng=es&nrm=iso&tlng=es

[14] "Instituto Nacional de Estadistica e Informatica." Accessed: Jun. 21, 2024. [Online]. Available: https://m.inei.gob.pe/prensa/noticias/el-33-del-gasto-en-alimentos-de-los-peruanos-son-realizados-fuera-del-hogar-8539/

[15] A. Marí-Sanchis, J. Burgos-Balmaseda, and R. Hidalgo-Borrajo, "Eating disorders in sport. Update and proposal for an integrated approach," Endocrinology, Diabetes and Nutrition (English ed.), vol. 69, no. 2, pp. 131–143, Feb. 2022, doi: 10.1016/J.ENDIEN.2022.02.016.

[16] "Nutrition in Universal Health Coverage." Accessed: Jun. 21, 2024. [Online]. Available: https://iris.who.int/handle/10665/331148

[17] B. Başpinar, N. N. Aslan Çin, A. Özfer Özçelik, B. Başpinar, N. N. Aslan Çin, and A. Özfer Özçelik, "Is Diet Quality In Individuals Who Engage In Regular Physical Activity Affected By Nutritional Knowledge And Practices?: A Cross-Sectional Study," Journal of the Faculty of Human Medicine, vol. 23, no. 2, pp. 38–48, Apr. 2023, doi: 10.25176/RFMH.V23I2.5638.

[18] S. Reyes Narvaez, M. O. Canto, S. Reyes Narvaez, and M. O. Canto, "Knowledge about healthy eating in students of a public university," Chilean Journal of Nutrition, vol. 47, no. 1, pp. 67–72, Feb. 2020, doi: 10.4067/S0717-75182020000100067.

[19] C. Wanden-Berghe and C. Wanden-Berghe, "Nutritional Assessment in the Elderly," Hospital a Domicilio, vol. 6, no. 3, pp. 121–134, Jul. 2022, doi: 10.22585/HOSPDOMIC. V6I3.171.

[20] D. J. Fuentes-Sapiencia and J. F. Fuentes-Carrasco, "Food and nutritional factors that optimize growth and development in a high-performance gymnast in a case study," Cuadernos Hospital de Clínicas, ISSN 1562-6776, Vol. 62, No. 2, 2021, pp. 52-56, vol. 62, no. 2, pp. 52–56, 2021, Accessed: Jun. 21, 2024. [Online]. Available: https://dialnet.unirioja.es/servlet/articulo?codigo=8964851&info=resumen&idioma=SPA

[21] Y. F. Márquez-Sandoval et al., "Design and validation of a questionnaire to assess eating behavior in Mexican students in the area of health," Nutr HospVol. 30 No. 1 pp. 153–164, 2014, doi: 10.3305/n.2014.30.1.7451.

[22] A. Izquierdo Hernández, M. Armenteros Borrell, L. Lancés Cotilla, and I. Martín González, "Alimentación saludable," Rev Cubana Enferm, vol. 20, no. 1, pp. 1–1, 2004, Accessed: Jun. 21, 2024. [Online]. Available: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-03192004000100012&lng=es&nrm=iso&tlng=es

[23] M. M. P. Campos, S. Ramos-Cosi, and L. Andrade-Arenas, "SAFE Mobile Application: Prevention of Violence Against Women," International Journal of Engineering Trends and Technology, vol. 71, no. 12, pp. 299–307, Dec. 2023, doi: 10.14445/22315381/IJETT-V71I12P228.

[24] "Profile of people who attend gyms in the city of Catamarca and its relationship with the reasons for practicing bodybuilding activities." Accessed: Jun. 21, 2024. [Online]. Available: https://www.efdeportes.com/efd118/perfil-de-las-personas-que-asisten-a-los-gimnasios.htm#google_vignette

[25] E. R. Helms, A. A. Aragon, and P. J. Fitschen, "Evidence-based recommendations for natural bodybuilding contest preparation: Nutrition and supplementation," J Int Soc Sports Nutr, vol. 11, no. 1, p. 20, May 2014, doi: 10.1186/1550-2783-11-20.

[26] "Salud - Desarrollo Sostenible." Accessed: Jun. 21, 2024. [Online]. Available: https://www.un.org/sustainabledevelopment/es/health/

[27] E. Militar de Cadets, G. José María Córdova, and C. Zafra Galvis, "Revista Científica General José María Córdova," General José María Córdova, vol. 4, pp. 13–14, 2006, Accessed: Jun. 21, 2024. [Online]. Available: http://www.redalyc.org/articulo.oa?id=476259067004

[28] And. Del Canto and A. S. Silva, "QUANTITATIVE METHODOLOGY: BOARDING FROM THE COMPLEMENTARITY IN THE SOCIAL SCIENCES KEYWORDS: RESEARCH * SOCIAL SCIENCES * RESEARCH METHOD * QUANTITATIVE ANALYSIS * QUALITATIVE ANALYSIS * COMPLEMENTARITY," Rev. Social Sciences, vol. 141.

[29] "Gimnasios en Perú - Gimnasios.com." Accessed: Jun. 21, 2024. [Online]. Available: https://www.gimnasios.com.pe/

[30] J. C. Siurana Aparisi, "The Principles of Bioethics and the Emergence of an Intercultural Bioethics," Veritas, vol. 22, no. 22, pp. 121–157, Mar. 2010, doi: 10.4067/S0718-92732010000100006.

[31] R. M. Sanadi, P. D. Khandekar, S. R. Chaudhari, M. A. Javali, and N. U. Gurav, "Association of periodontal disease with oral lichen planus: A systematic review and meta analysis," Journal of Oral and Maxillofacial PathologyVol. 27 No. 1 pp. 173–180, Jan. 2023, doi:10.4103/jonfp.jonfp_178_22.

[32] K. A. Pituch and J. P. Stevens, Applied Multivariate Statistics for the Social Sciences: Analyses with SAS and IBM's SPSS, Sixth Edition. Taylor and Francis, 2015. doi: 10.4324/9781315814919.

[33] D. Klu, D. Y. Atiglo, and A. K. Christian, "Multinomial logistic regression analysis of the determinants of anaemia severity among children aged 6–59 months in Ghana: new evidence from the 2019 Malaria Indicator Survey," Volume 23, Issue 1, Vol. 23 No. 1 John. 1991, doi:10.1186/C12887-023-03919-0.

[34] World Health Organization, "Ciclo de Vida." Accessed: Aug. 14, 2024. [Online]. Available: https://www.minsalud.gov.co/proteccionsocial/Paginas/cicloVida.aspx

[35] A. A. Huamani-Huaracca, M. N. Zela-Garcia, M. M. Rojas-Villarreal, F. D. M. Quinteros-Domínguez, A. M. Flores-Hiyo, and J. Morales, "Muscle Dysmorphia and Satisfaction with Body Image among Gym Users in Lima," Volume 12, Issue 5, Pages 759 - 768, vol. 12, no. 5, pp. 759–768, Sep. 2024, doi: 10.13189/saj.2024.120501.

Development of a Causal Model of Post-Millennials' Willingness to Disclose Information to Online Fashion Businesses (Thailand)

Apiwat Krommuang[1], Jinnawat Kasisuwan[2]
KMITL Business School, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Thailand[1, 2]

***Abstract*—This research examines the causal factors influencing the willingness of Central Post-Millennials to disclose information to online fashion businesses by using privacy calculus theory as the basic principle for modeling. The study has three primary objectives: (1) to investigate the causal factors influencing willingness to disclose information, (2) to analyze both the direct and indirect effects of perceived risk, perceived benefit, perceived value, perceived control over the use of personalization data, and trust on the willingness to disclose information, and (3) to develop a causal factor model for understanding the determinants of willingness to disclose information among Central Post-Millennials in the context of online fashion businesses. The research sample consists of 385 individuals, and data were collected using a structured questionnaire. Descriptive and inferential statistical methods were employed for data analysis. The relationships between variables were assessed using Pearson's Correlation Coefficient. The model's fit to the empirical data was evaluated using goodness-of-fit measures, and the transmission of influence was tested through structural equation modeling (SEM). The findings reveal that demographic factors do not significantly affect the willingness to disclose information. However, the study identifies perceived risk, perceived benefit, perceived value, perceived control over the use of personalization data, and trust as key determinants of willingness to disclose information to online fashion businesses. Among these, perceived control exhibits the strongest influence, closely followed by trust. These results highlight the antecedent processes influencing the willingness to disclose information, as represented by a model developed from a comprehensive literature review and empirically tested for consistency with the data.**

Keywords*—Online fashion business; Post-Millennials; privacy calculus; willingness; disclose information

I. Introduction

In the contemporary era of rapid technological and digital advancement, consumers have gained unprecedented access to the internet. This proliferation serves as a clear indicator that the Thai populace has embraced and integrated technology into their quotidian existence. Statistical data demonstrates that there was an average annual increase of 3.6 million internet users in Thailand from 2013 to 2021 [1].

Further analysis of internet usage patterns reveals that Generation Y exhibits the highest average daily usage, closely followed by Generation Z, with the latter being the most prolific users overall [1]. The term "Post Millennials" is employed to denote individuals born from 1997 to the present, encompassing both late Generation Y and Generation Z cohorts. This demographic represents one of the largest generational groups entering the labor market to date [2]. Notably, the Post-Millennial cohort demonstrates the highest propensity for online purchasing across all generations. Within this context, the most prevalent product categories acquired through e-commerce channels are apparel, footwear, sporting equipment, and accessories, which consistently rank as the primary items of online consumption.

The expansion of the internet user base presents a significant opportunity for the development of digital enterprises and the establishment of organizations adept at navigating the digital landscape. Consequently, there is an influx of consumer data into the digital realm. From a commercial perspective, data represents a valuable resource and plays a pivotal role in business development across all sectors. Both small and large organizations are currently engaged in the process of leveraging data to create maximal value. Entities that employ the data-driven organization paradigm utilize data as the foundation for their strategic decisions. A survey conducted by PricewaterhouseCoopers [4], encompassing over 1,000 senior executives, revealed that highly data-driven organizations tend to exhibit greater efficacy in decision-making processes and demonstrate threefold superior performance compared to their counterparts who rely less on data [3].

However, as the reliance on consumer data intensifies, concerns about privacy have surged. The unauthorized sharing of sensitive consumer information by businesses and governmental entities worldwide reveals a critical issue regarding personal data security. This encompasses not only data breaches and potential security threats but also a significant erosion of user autonomy over their personal information. [5]

In light of these opportunities and challenges, it has been observed that technological advancements facilitating expedited and convenient access to various services through online channels, in exchange for personal information, are considered the genesis of privacy erosion. This phenomenon aligns with the Privacy Calculus theory, a seminal concept in privacy research, which emphasizes individual decision-making regarding the disclosure of personal information based on a cost-benefit analysis of potential risks and benefits [6].

Consequently, the researcher is compelled to employ the Privacy Calculus theory as a theoretical framework to investigate the factors influencing the willingness to disclose personal information. The Central Post-Millennials cohort, characterized by their substantial online fashion product

purchases, presents an appropriate demographic for this study. The findings of this research will serve as a foundation for strategic planning and adaptation to address evolving consumer needs with greater precision while simultaneously enhancing user confidence in online fashion purchasing systems.

The remainder of this document will delve into the strategies for collecting consumer data effectively, which is critical for transforming the organization into a data-driven enterprise. This discussion aims to offer insights into aligning business goals with consumer behavior in the digital era.

II. Literature Review

A. Privacy Calculus

The concept of Privacy Calculus, originally developed by Laufer and Wolfe [7], has significantly contributed to the understanding of data privacy, user behavior, and decision-making processes within the digital environment. Numerous researchers have utilized this theory to develop models that predict and explain users' privacy-related actions. The theory enhances transparency in consumer data collection practices by offering clear explanations of the benefits consumers will receive. It is evident that the application of the Privacy Calculus concept can be integrated into various research frameworks, leading to numerous studies that utilize this concept to examine privacy in diverse contexts, depending on the specific objectives of those studies. In this research, the Privacy Calculus concept encompasses Perceived Risk, Perceived Benefits, Perceived Value of Information Disclosed, Perceived Control over the Use of Personalization Data, Trust, and Willingness to Disclose Information.

Perceived Risk (PR) relates to the uncertainty stemming from the potential opportunistic behavior of service providers, which may result in consumer losses [8]. From the Privacy Calculus perspective, this encompasses the perception of potential loss from information disclosure [9]. The perception of risk, as understood by consumers, is underpinned by Prospect Theory, developed by [42] which addresses decision-making under risk [10]. This theory posits that decisions often prioritize the avoidance of losses over the pursuit of gains. Accordingly, this research employs the variable of risk perception to measure the extent of consumer concern regarding the security of personal information.

Previous studies, such as [11] have demonstrated that risk perception negatively impacts the willingness to disclose health information among patients in China, consistent with other studies including [12] who examined the specific risks and benefits influencing consumer acceptance of mobile location-based advertising (MLBA); [13] who explored the interdependence of personality contingency calculus and causal asymmetry in social networking site (SNS) information disclosure behavior; [14] who analyzed cloud storage users' intentions to include personal information in cloud storage applications in Indonesia and Taiwan; [15] who investigated the privacy paradox related to information disclosure in mHealth applications; and [16] who studied the effect of trust on students' willingness to use online learning platforms.

Hypothesis 2. Perceived Risk negatively influences the willingness of Central Post-Millennials to disclose information to online fashion businesses.

Moreover, [17] explored the key determinants of experts' attitudes and intentions to use VC apps among 484 surveyed experts. Their study found that perceived risk influenced the perceived value of information disclosure, consistent with other studies such as [18] who examined factors affecting personal data disclosure among mobile application users in Bangkok; [19] who explored the relationship between benefits, trust, and honesty in personal data disclosure within mobile banking services in Vietnam; and [20] who studied the key determinants of contact tracing app adoption in Australia.

Hypothesis 3. Perceived Risk negatively influences the Perceived Value of Information Disclosed by Central Post-Millennials to online fashion businesses.

Perceived Benefits (PB) refer to the perceived advantages obtained from disclosing information to access a service. This concept involves balancing the need for privacy with the need for disclosure to attain certain benefits [21]. Perceived benefits can be categorized into monetary and social rewards [22] and other factors such as performance expectations [20]; personal interests [14]; and extrinsic rewards [23]. Expectancy-Value Theory, developed by Atkinson [24], supports the perception of potential benefits by evaluating decision-making processes based on expected outcomes and the satisfaction derived from them. From the Privacy Calculus perspective, this theory suggests that the satisfaction from potential outcomes of information disclosure influences individuals' expectations and decisions, leading to a higher likelihood of information disclosure if the perceived benefits outweigh the risks. This research, therefore, employs the perceived benefits variable to measure the extent to which consumers perceive the value or benefits of disclosing personal information.

Previous research by [25], which examined the motivating and inhibiting factors affecting customers' willingness to disclose personal data on e-commerce websites in Saudi Arabia, indicated a significant positive relationship between perceived benefits and willingness to disclose personal data. Similar findings were reported by [11] regarding health information disclosure in China; [26] on the rational and irrational factors influencing data disclosure decisions; [23] on the motivation behind Internet users' willingness to provide personal data; and [12] on the acceptance of MLBA. Additionally, studies by [13, 14, 15, 16] have provided consistent findings with those of [25].

Hypothesis 4. Perceived Benefits positively influence the willingness of Central Post-Millennials to disclose information to online fashion businesses.

Numerous studies have supported the relationship between perceived benefits and the value perspective of information disclosure, alongside perceived risks, as highlighted by [27], who studied factors affecting consent to disclose personal data among commercial bank application users in Bangkok. Research by [18, 19, 20] also confirmed this relationship.

Hypothesis 5. Perceived Benefits positively influence the Perceived Value of Information Disclosed by Central Post-Millennials to online fashion businesses.

Perceived Value of Information Disclosed (PV) refers to the value consumers attribute to the personal information they disclose to service providers, derived from a comparison of the benefits gained against the risks incurred. This assessment can be understood through a Cost-Benefit Analysis framework [6], where the perceived value is determined by evaluating the trade-offs between perceived benefits and risks. The theory of Customer-Perceived Value (CPV) by [28] further elaborates that this value arises from the difference between customers' evaluations of all benefits and the total costs associated with an offer. These benefits and costs include various dimensions, such as financial, time, energy, and psychological aspects.

The relationship between perceived value of information disclosed and the willingness to disclose information has been supported by several studies, including [29] who studied the factors influencing the privacy calculus of wearable fitness devices among Korean undergraduate students and found a significant positive impact of perceived value on information disclosure intentions. Other studies, such as [16, 30, 18, 27] have also confirmed this relationship.

Hypothesis 6. Perceived Value of Information Disclosed positively influences the willingness of Central Post-Millennials to disclose information to online fashion businesses.

Perceived Control over the Use of Personalization Data (PC) refers to an individual's belief in their ability to manage the disclosure and dissemination of their personal data [9]. This perception of control can be supported by Control Balance Theory, developed by [31], which emphasizes the balance of control between individuals and systems in societal contexts. In the context of Privacy Calculus, perceived control pertains to the extent to which individuals believe they can influence how their personal data is used. An imbalance in this control, where users feel they have less power over their data compared to data controllers, can lead to decreased trust and willingness to disclose information.

Past research by [19], which examined personal data disclosure in mobile banking services in Vietnam, supported the positive relationship between perceived data control and the value perspective of data disclosure. Additionally, [32] explored the influence of perceived control on the tension between privacy concerns and data disclosure among Chinese users, finding a significant positive effect on self-disclosure intentions.

Hypothesis 7. Perceived Control over the Use of Personalization Data positively influences the willingness of Central Post-Millennials to disclose information to online fashion businesses.

Hypothesis 8. Perceived Control over the Use of Personalization Data positively influences the Perceived Value of Information Disclosed by Central Post-Millennials to online fashion businesses.

Trust (T) refers to consumers' confidence in the proper management, reliability, and security of their personal information when shared with service providers through websites or the Internet [6]. Trust can be examined through various dimensions, including Interpersonal Trust, System Trust, and Organizational Trust [33]. This research focuses on trust in organizations and systems, commonly utilized within the Privacy Calculus framework [6, 34, 35, 36].

Previous studies, such as [37], which explored customers' willingness to disclose personal information on small B2C e-commerce websites, have found a positive relationship between trust and willingness to disclose information. This finding is consistent with research by [14, 27, 18].

Hypothesis 9. Trust positively influences the willingness of Central Post-Millennials to disclose information to online fashion businesses.

Willingness to Disclose Information (WDI) refers to an individual's acceptance and understanding of the expectations related to personal information disclosure, reflecting a readiness to share information based on situational needs. This concept is often related to the Willingness to Pay theory, where the price a consumer is willing to pay is determined by their assessment of a product or service's value [38]. In the context of privacy, this concept can be interpreted as the risk associated with disclosing information, where consumers evaluate the trade-off between the risks and benefits of sharing personal data. Consequently, this research uses this variable to measure the level of consumer willingness to disclose personal information to online fashion businesses.

B. Model Conceptualization

Path diagrams offer significant advantages in research by clearly illustrating causal relationships between variables and supporting both direct and indirect relationship analysis. They provide structured, easy-to-interpret visual representations, making complex theoretical frameworks more accessible.

The conceptual framework of this research is constructed based on the examination of relevant theories and prior studies. It identifies the causal factors anticipated to influence the perceived value of information disclosure, which include Perceived Risk, Perceived Benefit, and Perceived Control over the Use of Personalization Data [11, 27, 17, 18, 19, 20]. Additionally, the factors expected to influence the willingness to disclose information are Perceived Risk, Perceived Benefit, Perceived Value of Information Disclosed, Perceived Control over the Use of Personalization Data, and Trust [25, 26, 27, 18, 23, 12, 32, 13, 29, 14, 15, 30, 16]. Accordingly, the research conceptual framework is depicted in Fig. 1.

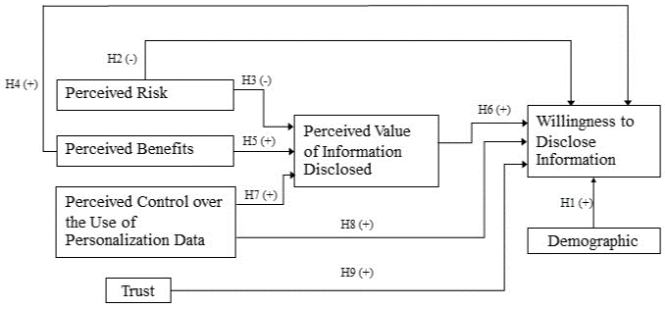


Fig. 1. Research conceptual framework.

III. METHODOLOGY AND RESULTS

The population targeted in this research includes individuals born between 1997 and 2006, aged 18-27 years, residing in

Thailand, and who have disclosed information to a fashion business at least once within the past six months. Given the unknown exact population size, the sample size was determined using Cochran's formula [39] with a 95% confidence level, resulting in a required sample of 385 respondents. The sampling was conducted via purposive and quota sampling methods through online channels, particularly in a large public Facebook group, ensuring the inclusion of users across all purchasing channels. The sample was divided equally among the seven main online fashion shopping platforms: Instagram, TikTok, Shopee, Lazada, Line, SHEIN, and brand websites, with 55 respondents per platform. The data collection tool was a questionnaire, vetted for quality by three experts and tested for reliability. The statistical methods used in this research include frequency, percentage, mean, standard deviation, one-way analysis of variance, Pearson's correlation coefficient, goodness of fit measures, and Structural Equation Modeling (SEM).

A. Results of General Data

The analysis of general demographic data for the 385 respondents revealed that the majority were female, aged between 23-25 years, holding a bachelor's degree or equivalent, predominantly students, with an income range of 20,000 - 39,999 baht. In terms of online fashion purchasing behavior, clothing/apparel emerged as the most frequently purchased product, with an average purchase frequency of 3-5 times per month and an average expenditure of 400-699 baht per purchase.

B. Results of Demographic Factors Hypothesis Testing

Hypothesis testing on the differences in willingness to disclose information to online fashion businesses, based on demographic factors (age, education, occupation, and income), using t-test and one-way ANOVA, indicated no significant differences in willingness to disclose information across these demographic factors (see Table I).

TABLE I. RESULTS OF THE HYPOTHESIS TESTING ON WILLINGNESS TO DISCLOSE INFORMATION CLASSIFIED BY DEMOGRAPHIC CHARACTERISTICS USING T-TEST AND ONE-WAY ANOVA STATISTICS

Demographic	The results of the hypothesis testing
Age	Not significantly
Education	Not significantly
Occupation	Not significantly
Income	Not significantly

C. Results of Correlation Coefficients Between Variables

The study of the relationships between five causal factors: Perceived Risk (PR), Perceived Benefit (PB), Perceived Value of Information Disclosed (PV), Perceived Control over Personal Data Use (PC), and Trust (T) and the willingness to disclose information (WDI) among Central Post-Millennials to online fashion businesses found that all variable pairs were significantly correlated, with correlation coefficients ranging from 0.152 to 0.527. These values meet the criterion of not exceeding a correlation coefficient of 0.85, indicating no multicollinearity issues [40]. Table II shows Pearson's correlation coefficient between variables.

TABLE II. PEARSON'S CORRELATION COEFFICIENT BETWEEN VARIABLES

	PR	PB	PV	PC	T	WDI
PR	1					
PB	-0.209***	1				
PV	-0.285***	0.525***	1			
PC	-0.448***	0.203***	0.152 **	1		
T	-0.281***	0.358***	0.160 **	0.185 ***	1	
WDI	-0.527***	0.426***	0.351 ***	0.518 ***	0.484 ***	1

Note : * Statistical significance level is 0.05
** Statistical significance level is 0.01
*** Statistical significance level is 0.001

D. Results of Goodness of Fit Assessments for the Model

The initial analysis of the model's consistency, as shown in Fig. 2, with the empirical data indicated that it met all criteria, as presented in Table III. However, the relationship between Perceived Control Over Personal Data Use (PC) and Perceived Value of Information Disclosed (PV) was statistically insignificant. This may be attributed to the absence of a genuine linear relationship between the two variables.

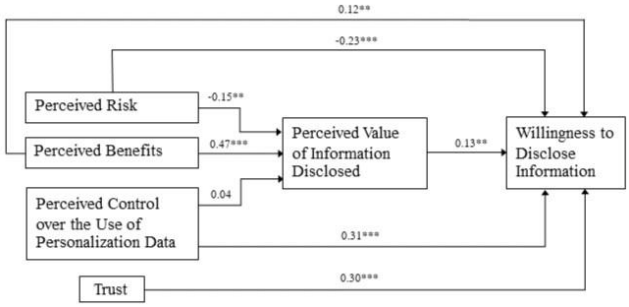


Fig. 2. Initial model.

Therefore, the model was adjusted by removing this relationship as shown in Fig. 3. The revised model continued to meet all criteria, confirming its consistency with the empirical data, with no statistically insignificant relationships remaining.

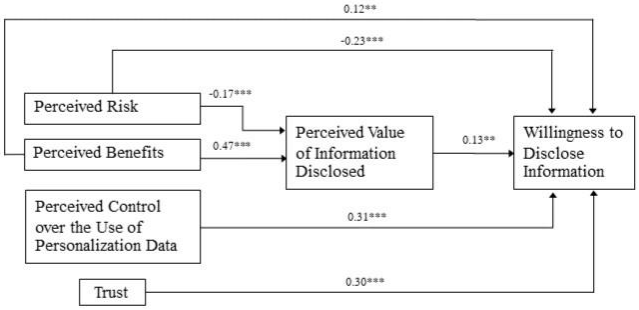


Fig. 3. Adjusted model.

TABLE IV. RESULTS OF THE MODEL'S FIT TO THE EMPIRICAL DATA WAS EVALUATED USING GOODNESS-OF-FIT MEASURES

Goodness of fit measures	Criteria	Initial model m	Adjusted model
P-value of Chi-square Statistics (χ^2)	>0.05	0.14	0.26
Chi-square/Degree of freedom (χ^2/df)	<3.00	2.17	1.35
Comparative Fit Index (CFI)	>0.95	0.9981	0.9988
Goodness of Fit Index (GFI)	>0.95	0.9981	0.9977
Normed Fit Index (NFI)	>0.95	0.9965	0.9956
Root Mean Squared Residuals (RMR)	<0.05	0.0051	0.0054

E. Hypothesis Testing of Relationship Path

The hypothesis testing using SEM to assess the causal factors affecting the willingness to disclose information to online fashion businesses produced the results summarized in Table IV. The findings indicate that hypotheses H2, H3, H5, H8, and H9 were accepted with a significance level of 0.001, and hypotheses H4 and H6 were accepted with a significance level of 0.01. However, Hypothesis H7 was rejected, as depicted in Fig. 2.

TABLE V. RESULTS OF HYPOTHESIS TESTING AMONG RELATIONSHIP PATH OF CAUSAL VARIABLES

Hypothesis	Relationship Path	Intensity	Result
H2	PR → WDI	-0.23***	Accept
H3	PR → PV	-0.17***	Accept
H4	PB → WDI	0.12**	Accept
H5	PB → PV	0.47***	Accept
H6	PV → WDI	0.13**	Accept
H7	PC → PV	-	Reject
H8	PC → WDI	0.31***	Accept
H9	T → WDI	0.30***	Accept

Note : * Statistical significance level is 0.05
** Statistical significance level is 0.01
*** Statistical significance level is 0.001

F. Results of the Intensity between the Causal Relationship

The analysis revealed that Perceived Risk and Perceived Benefits have direct influences on the Perceived Value of Information Disclosed, with path coefficients of -0.17 and 0.47, respectively. Additionally, Perceived Risk, Perceived Benefits, Perceived Control over Personal Data Use, Trust, and Perceived Value of Information Disclosed directly influence the willingness to disclose information, with path coefficients of -0.23, 0.12, 0.31, 0.30, and 0.13, respectively. Perceived Risk and Perceived Benefits also have indirect influences on the willingness to disclose information through the Perceived Value of Information Disclosed, with coefficients of -0.02 and 0.03, respectively.

The total effects of the causal factors Perceived Risk, Perceived Benefits, Perceived Value of Information Disclosed, Perceived Control over Personal Data Use, and Trust on the willingness to disclose information to online fashion businesses were found to be -0.25, 0.18, 0.13, 0.31, and 0.30, respectively.

The model's squared multiple correlation (R^2) indicates that the variables can explain 30% of the variance in the Perceived Value of Information Disclosed and 53% of the variance in the willingness to disclose information. Table V shows results of the analysis of DE, IE, TE between variables.

TABLE VI. RESULTS OF THE ANALYSIS OF DIRECT INFLUENCE (DE), INDIRECT INFLUENCE (IE), TOTAL INFLUENCE (TE) BETWEEN VARIABLES AND THE SQUARED MULTIPLE CORRELATION COEFFICIENT (R-SQUARE : R^2)

	PV			WDI		
	DE	**IE**	**TE**	**DE**	**IE**	**TE**
PR	-0.17***	-	-0.17***	-0.23***	-0.02**	-0.25***
PB	0.47***	-	0.47***	0.12**	0.03**	0.18***
PC	-	-	-	0.31***	-	0.31***
T	-	-	-	0.30***	-	0.30***
PV	-	-	-	0.13**	-	0.13**
R^2	0.30			0.53		

IV. DISCUSSIONS

A. Discussion of Findings

The findings reveal that respondents with different demographic characteristics did not show significant differences in their willingness to disclose information to online fashion businesses (rejecting H1). This outcome may be attributed to the pervasive involvement of consumers in the online world and social media in the current era. Given that internet usage and awareness of information security have become integral aspects of post-millennial consumers' daily lives, their level of privacy awareness is relatively uniform. Additionally, the widespread education and dissemination of knowledge regarding risks and personal data protection have contributed to this trend. This aligns with the findings of [41], which indicated that control variables (e.g., age, income, education, gender, fear, privacy violation) did not significantly affect respondents' intention to disclose information to websites. This suggests that the intention to disclose information is more effectively explained by the independent variables in the research model rather than by demographic differences among respondents.

The causal factor of Perceived Risk influences the willingness of Central Post-Millennials to disclose information to online fashion businesses, primarily due to uncertainties and concerns about potential impacts. Uncertainty regarding data security arises from the perception that providing information online may lead to unexpected problems, as well as fears of unauthorized data use. Users who believe their information may be misused without permission often feel insecure and worried. The risk of third-party access to their information also contributes to discomfort. When users are aware of these risks, they are more likely to withhold information or take additional protective measures, prioritizing privacy protection. This perceived risk exerts a negative direct influence on willingness to disclose information (-0.23, accepting H2) and an indirect influence through the perceived value perspective of disclosure (-0.02). This indicates that concerns about risk lead consumers

to be more cautious and less inclined to disclose information. When consumers perceive higher risks, their willingness to disclose information decreases, which is consistent with [14], who found that cloud storage users' willingness to disclose personal information is significantly influenced by perceived costs. Similarly, [20] demonstrated that privacy risk perception significantly influences contact tracing app usage intentions through the mediating effect of information disclosure value.

The causal factor of perceived benefits also affects the willingness of Central Post-Millennials to disclose information to online fashion businesses. Benefits such as discounts attract and motivate users to share personal information. Additionally, users satisfied with service providers' reuse of their personal data to offer personalized services are more inclined to disclose information, viewing personalized services as valuable benefits. Central Post-Millennials, a significant group in the online fashion market, are driven by their pursuit of distinctive experiences and personalized shopping. This factor has a direct positive influence on willingness to disclose information (0.12, accepting H4) and an indirect influence through the perceived value perspective of disclosure (0.03). Although the positive influence is not as substantial as other factors, it demonstrates that consumers value the benefits of disclosing information, motivating them to share more willingly. This finding is consistent with [20] who found that performance expectation benefits significantly influence contact tracing app usage intentions through the perceived value of disclosure. The study in [32] also confirmed that perceived benefits significantly affect self-disclosure intentions.

The perceived value of disclosure significantly influences the willingness of Central Post-Millennials to disclose information to online fashion businesses. Consumers often compare risks and benefits, and when they perceive that the risks are acceptable relative to the benefits, they are more inclined to disclose information. Users who view the benefits as outweighing the risks feel that disclosing information is a valuable and rational decision. When users believe that disclosing information is worthwhile, they exhibit a greater willingness to share personal information. This perception of value promotes a positive attitude toward information disclosure, even when some risks are present. It shows that users evaluate the value and benefits carefully and make rational decisions, resulting in a direct positive influence on willingness to disclose information (0.13, accepting H6) with no indirect influence. This factor has the least overall influence on willingness to disclose information among all factors, indicating that while consumers make rational comparisons, the perceived value is not as impactful as other factors. This is consistent with the findings of [18], who observed that the perceived value of disclosing information significantly affects the willingness to disclose information.

The perceived control over personal data use is another significant factor affecting the willingness of Central Post-Millennials to disclose information to online fashion businesses. Users who feel they have control over their personal data are more confident and comfortable sharing it. The ability to decide on their data makes users feel secure and empowered to protect their data. Additionally, users who understand the policies and methods of data usage are more confident in a transparent system. Providing clear information about personal data use fosters a sense of involvement in decision-making, increasing the likelihood of information disclosure. When users perceive that their privacy is respected and their rights are protected, they feel more empowered to share their data. This factor has a direct positive influence on willingness to disclose information (0.31, accepting H7) and no indirect influence, making it the most influential factor in promoting willingness to disclose information, even without an indirect effect. This is consistent with [32] who found that perceived controllability significantly impacts self-disclosure intentions.

Trust is another crucial factor influencing the willingness of Central Post-Millennials to disclose information to global fashion businesses. Users who feel confident in a brand's or seller's ability to safeguard their information are more willing to disclose it, believing their personal information will not be misused. Trust in the platform's policies and protective measures enhances confidence in sharing information. Clear policies and strict protection measures help build trust and encourage more information disclosure. Positive online shopping experiences also contribute to increased willingness to disclose information, as users perceive that sharing data is safe and properly protected. Trust fosters a positive relationship between users and platforms, further increasing the willingness to disclose information. Trust has a direct positive effect on willingness to disclose information (0.30, accepting H9) and no indirect effect. This influence is comparable to perceived control, which is the most significant factor affecting willingness to disclose information. Trust has the second-highest positive impact on willingness to disclose information, following perceived control. This aligns with the findings of [14], who indicated that cloud storage users' willingness to share personal data is significantly influenced by trust, perceived costs, perceived benefits, and the sensitivity of personal data.

V. CONCLUSION

The development of a causal model to understand the factors influencing the willingness of Central Post-Millennials to disclose information to online fashion businesses involves a comprehensive literature review to identify key determinants. The factors identified as likely to affect the willingness to disclose information include Perceived Risk, Perceived Benefits, Perceived Value of Information Disclosed, Perceived Control over Personal Data Use, and Trust. These factors were then integrated into a path model based on relationships supported by the literature. After conducting statistical tests, the model fit indices indicated that the revised model accurately and reliably explains the relationships between these factors, thereby elucidating the behavior of Central Post-Millennial consumers. The multiple correlation coefficient (R-SQUARE: R^2) for predicting the willingness to disclose personal data was found to be 0.53 or 53%.

Given the intensity of influence, businesses should prioritize strategies that enhance Perceived User Control over Personal Data Use and trust, as these factors exert the strongest influence on the willingness to disclose information. Strategies could focus on creating transparency that simultaneously addresses both factors, reinforcing consumer confidence and control. Furthermore, while the Perceived Value of disclosure reflects

that consumers do consider a risk-benefit comparison, it is perceived as cumbersome, resulting in the lowest influence among the factors affecting willingness to disclose information. However, this process shouldn't be disregarded or underestimated, as it acts as a mediator, indirectly passing the influence of both perceived risks and perceived benefits on the dependent variable.

Emphasizing Perceived User Control over Personal Data Use and Trust can lead policymakers to advocate for regulations requiring businesses to provide clear information on data practices. Strengthening consumer rights to manage personal data, including consent options, aligns with users' desires for control. By prioritizing these elements in legislation, policymakers can enhance consumer confidence and promote responsible data practices, fostering a more secure digital marketplace. Additionally, key insights for marketing strategies. Marketers can enhance transparency about data practices, boosting consumer confidence and encouraging information disclosure. Empowering consumers with data management options fosters stronger relationships and brand loyalty, helping businesses comply with regulations while building trust in a privacy-conscious marketplace.

REFERENCES

[1] Electronic Transactions Development Agency. (2022). Thailand Internet User Behavior 2022. [Online] Available at: https://www.etda.or.th/

[2] Ozkan, M., and Solmaz, B. (2015). The Changing Face of the Employees – Generation Z and Their Perceptions of Work (A Study Applied to University Students). Procedia Economics and Finance, (1), 476-483.

[3] Stobierski, T. (2021). The advantages of data-driven decision-making. Harvard Business School Online. [Online] Available at: https://online.hbs.edu/blog/post/data-driven-decision-making.

[4] PricewaterhouseCoopers. 2023. PwC Global Consumer Insights Survey 2023. [Online] PwC. Available at: https://www.pwc.com/gx/en/

[5] Aras, A., & Tekin, K. (2022). Privacy concerns in consumer E-commerce activities and response to social media advertising: Empirical evidence from Europe. Computers in Human Behavior, 128, 107086.Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. Information Systems Research, 17(1), 61-80.

[6] Laufer, R., & Wolfe, M. (1977). Privacy as a Concept and a Social Issue: A Multidimensional Developmental Theory. Journal of Social Issues, 33(3), 22-42.

[7] Ganesan, S. (1994) Determinants of Long-Term Orientation in Buyer-Seller Relationships. The Journal of Marketing, 58, 1-19.

[8] Smith, H. J., Dinev, T., & Xu, H. (2011). Information Privacy Research: An Interdisciplinary Review. MIS Quarterly, 35(4), 989-1015.

[9] Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. Econometrica, 47(2), 263-292.

[10] Shi, J., Yuan, R., Yan, X., Wang, M., Qiu, J., Ji, X., & Yu, G. (2023). Factors influencing the sharing of personal health data based on the integrated theory of privacy calculus and theory of planned behaviors framework: Results of a cross-sectional study of Chinese patients in the Yangtze River Delta. Journal of Medical Internet Research, 25(7), e46562.

[11] Gutierrez, A., O'Leary, S., Rana, N. P., Dwivedi, Y. K., & Calle, T. (2019). Using privacy calculus theory to explore entrepreneurial directions in mobile location-based advertising: Identifying intrusiveness as the critical risk factor. Computers in Human Behavior, 95, 295-306.

[12] Sun, Y., Wang, N., & Shen, X.-L. (2021). Calculus interdependency, personality contingency, and causal asymmetry: Toward a configurational privacy calculus model of information disclosure. Information & Management, 58(8), 103556.

[13] Widjaja, A. E., Chen, J. V., Sukoco, B. M., & Ha, Q.-A. (2019). Understanding users' willingness to put their personal information on the personal cloud-based storage applications: An empirical study. Computers in Human Behavior, 91, 167-185..

[14] Zhu, M., Wu, C., Huang, S., Zheng, K., Young, S. D., Yan, X., & Yuan, Q. (2021). Privacy paradox in mHealth applications, an integrated elaboration likelihood model incorporating privacy calculus and privacy fatigue. Telematics and Informatics, 61, 101601

[15] Jiang X, Goh T-T and Liu M (2022). On Students' Willingness to Use Online Learning: A Privacy Calculus Theory Approach. Front. Psychol. 13:880261.

[16] Sandhu, R. K., Vasconcelos-Gomes, J., Thomas, M. A., & Oliveira, T. (2023). Unfolding the popularity of video conferencing apps – A privacy calculus perspective. International Journal of Information Management, 68, 102569.

[17] Ruangpannitchayakul, S., Intuwonges, S., Sukhaparamate, S., Jantarakolica, T., Jantarakolica, K., & Teekasap, P. (2022). Factors Affecting Personal Information Disclosure of Online Shopping Application Users in Bangkok and Metropolitan Area. Journal of Interdisciplinary Research: Graduate Studies, 11(2), 132-145. https://doi.org/10.14456/jirgs.2022.17

[18] Bui Thanh Khoa. (2021). The Impact of the Personal Data Disclosure's Tradeoff on the Trust and Attitude Loyalty in Mobile Banking Services. Journal of Promotion Management, 27(4), 585-608.

[19] Duan, S.X. and Deng, H. (2021). Hybrid analysis for understanding contact tracing apps adoption. Industrial Management & Data Systems, 121, 1599–1616.

[20] Westin, A.F. (2003). Social and Political Dimensions of Privacy. Journal of Social Issues, 59, 431-453.

[21] Wang, D., Liu, T., & Shi, J. (2017). Development of Monetary and Social Reward Processes. Scientific Reports, 7, 11128.

[22] Yeh, C.-H., Wang, Y.-S., Lin, S.-J., Tseng, T. H., Lin, H.-H., Shih, Y.-W., & Lai, Y.-H. (2018). What drives internet users' willingness to provide personal information? Online Information Review, 42(6), 923–939.

[23] Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. Psychology Review, 64, 359-372.

[24] Al-Jabri, I. M., Eid, M. I., & Abed, A. (2020). The willingness to disclose personal information: trade-off between privacy concerns and benefit. Information & Computer Security, 28(2), 161-181.

[25] Fernandes, T., and Pereira, N. (2021). Revisiting the privacy calculus: Why are consumers (really) willing to disclose personal data online? Telematics and Informatics, 65, 101717.

[26] Sumoolwhech, N. ., & Sukkaew, D. (2023). Factors affecting consent to disclose personal information of commercial banking application users in Bangkok Metropolitan region. RMUTP Journal of Business and Innovation Management, 2(1), 8–22.

[27] Kotler, P., & Keller, K. L. (2016). Marketing Management. (15th global edition). Edinburgh: Pearson Education.

[28] Cho, J. Y., Ko, D., & Lee, B. G. (2018). Strategic approach to privacy calculus of wearable device user regarding information disclosure and continuance intention. KSII Transactions on Internet and Information Systems (TIIS), 12, 3356–3374.

[29] Koester, N., Cichy, P., Antons, D., et al. (2022). Perceived privacy risk in the Internet of Things: determinants, consequences, and contingencies in the case of connected cars. Electronic Markets, 32, 2333–2355.Tittle, 1995.

[30] Tittle, C. R. (1995). Control balance: Toward a general theory of deviance. Westview Press.

[31] Ma, X., Qin, Y., Chen, Z., & Cho, H. (2021). Perceived ephemerality, privacy calculus, and the privacy settings of an ephemeral social media site. Computers in Human Behavior, 124, 106928.

[32] Luhmann, N. (1979). Trust and power. New York: John Wiley & Sons.

[33] Kehr, F., Wentzel, D., & Mayer, P. (2013). Rethinking the privacy calculus: On the role of dispositional factors and affect. International Conference on Information Systems.

[34] Krasnova, H., & Veltri, N. F. (2010). Privacy Calculus on Social Networking Sites: Explorative Evidence from Germany and USA. Hawaii International Conference on System Sciences.

[35] Krasnova, H., Veltri, N. F., & Günther, O. (2012). Self-disclosure and Privacy Calculus on Social Networking Sites: The Role of Culture - Intercultural Dynamics of Privacy Calculus. Business & Information Systems Engineering, 4(3).

[36] Widjaja, A. E., Capistrano, E. P., & Ha, Q. A. (2021). The Effect of Trust Seals, Information Disclosure, and Gender on Customers' Trust and Willingness to Disclose Personal Information on Small B2C E-Commerce Website. In ICEBE 2020: Proceedings of the First International Conference of Economics, Business & Entrepreneurship, ICEBE 2020, 1st October 2020, Tangerang, Indonesia (p.175). European Alliance for Innovation.

[37] Phradon Pridasak (2023). Principles of Microeconomics. Thammasat University.

[38] Cochran, W. G. (1953). Sampling Techniques. New York: John Wiley & Sons.

[39] Kline, R. B. (2015). Principles and practice of structural equation modeling. Guilford publications.

[40] Wakefield, R. (2013). The influence of user affect in online information disclosure. Journal of Strategic Information Systems, 22(2), 157–174.

[41] Kahneman, D. and Tversky, A., 1979. Prospect Theory: An Analysis of Decision under Risk. Econometrica, 47(2), 263-292.

Facial Expression Classification System Using Stacked CNN

Aditya Wikan Mahastama[1], Edwin Mahendra[2], Antonius Rachmat Chrismanto[3]*,
Maria Nila Anggia Rini[4], Andhika Galuh Prabawati[5]
Faculty of Information Technology, Universitas Kristen Duta Wacana, Yogyakarta, Indonesia

***Abstract*—Automatic emotion recognition technology through facial expressions has broad potential, ranging from human-computer interaction to stress detection and blood pressure assessment. Facial expressions exhibit patterns and characteristics that can be identified and analyzed by image processing and machine learning methods. These methods provide a basis for the development of emotion recognition systems. This research develops a facial emotion recognition model using Convolutional Neural Network (CNN) architecture, a popular architecture in image classification, segmentation, and object detection. CNNs offer automatic feature extraction and complex pattern recognition advantages on image data. This research uses three types of datasets, FER2013, CK+, and IMED, to optimize the deep learning approach. The developed model achieved an overall accuracy of 71% on the three datasets combined, with an average precision, recall, and F1-Score of 71%. The results show that CNN architecture performed well in facial emotion classification, supporting potential practical applications in various fields.**

***Keywords*—FER; CNN; deep learning; image classification**

I. Introduction

Artificial intelligence and machine learning technologies have created new possibilities in interpreting human emotions through facial expressions. Facial expressions are one of the natural ways humans communicate emotions. Automatic emotion recognition technology, mainly through facial expressions, has broad potential in various applications, ranging from human-computer interaction to stress detection and blood pressure assessment [1]. Facial expressions, as one of the non-verbal communication mediums, exhibit patterns and characteristics that can be identified and analyzed by image processing and machine learning methods. This provides a basis for developing systems to interpret facial expressions to recognize human emotions [2].

Convolutional Neural Network (CNN) is a deep learning network introduced in the 1960s. CNN is also applied in computer vision and is generally used in image classification, segmentation, object detection, and video processing [3]. Supporting the implementation of FER, deep learning-based technology with Convolutional Neural Networks (CNN) architecture was used by previous researchers as a potential solution to overcome problems in face and expression recognition and classification. Convolutional Neural Networks (CNN) are increasingly used in FER because they automatically extract features from images [4].

Facial Emotion Recognition refers to the ability to identify and recognize emotions expressed through the human face, belonging to an important research topic in computer vision and artificial intelligence. Facial emotions play an essential role in human communication, helping to understand the intentions and feelings of others, with two-thirds of human communication conveyed through nonverbal components, of which facial expressions play a major role. FER has two main approaches: the traditional approach and the Convolutional Neural Network (CNN) based approach. The conventional approach involves face component detection, feature extraction, and expression classification. The initial stage of the CNN architecture involves taking an image as input, followed by a convolution process to extract essential features, such as edges, texture, and shape. Afterward, a subsampling or pooling process is applied to reduce the dimensionality of the feature map while retaining important information. After multiple convolution and pooling layers, the feature map is flattened into a one-dimensional vector and connected to the fully connected layer, which finally uses a Softmax function to convert the output score into a probability distribution over the existing classes [5].

Convolutional Neural Networks (CNN) are one of the Deep Learning models that consist of automatic feature extractors and trainable classifiers. CNNs are designed to understand high-dimensional complex data with a specialized architecture that integrates convolution and subsampling layers. Although many CNN architectures have been developed for various tasks, such as object and handwriting recognition, the basic principle of CNN is to achieve the best performance in pattern recognition [6]. CNN also has a large representation capacity, where it learns the best features at each layer of the visual hierarchy. This makes CNN effective in various computer vision problems, such as object and handwriting recognition. One of the main advantages of CNN is its weight-sharing concept, which reduces the number of parameters that need to be trained and improves generalization [7].

This research aims to address the challenges of facial emotion recognition (FER) by optimizing CNN architectures to enhance accuracy and efficiency. The motivation lies in creating a system that can accurately interpret emotions, particularly for real-world applications such as mental health monitoring and interactive AI systems. The proposed method reduces the need for manual feature extraction and offers scalability across diverse datasets, making it applicable in various fields, including stress detection and user experience enhancement. A CNN-based approach is proposed to address the FER problem, with empirical evaluation showing that the performance improves when the outputs of different structured CNNs are combined and averaged, compared to using a single CNN architecture.

*Corresponding author

This article is written as follows: first, the introduction includes the research background, the research objectives to be addressed, and the general method proposed. Section II contains related works that reference previous studies, their relations, advantages, and limitations, and also theoretical foundation of the methods. Section III is the methodology, followed by the results and discussion in Section IV, which comprehensively presents the research findings and its analysis. Finally, this article gives the conclusion and suggestions for further research in Section V.

II. Related Works

Liu, Zhang, and Pan developed a Facial Emotion Recognition model using Convolutional Neural Networks (CNN) with the FER2013 dataset. The architecture design, layer depth, and number of neurons greatly influence the model's effectiveness. Large-scale CNNs face overfitting challenges and require high computational power. The model consists of three subnets with 8 to 10 layers, including an input layer, and three convolution layers with 3x3 filters at 64, 128, and 256 filters, respectively, followed by a max-pooling layer. Furthermore, there are three fully connected layers with 4096, 4096, and 7 neurons, ending with a SoftmaxLoss layer for classification. The third subnet performed best with 65.03% validation accuracy, especially in the surprise emotion category. However, the accuracy of the training data was not specified, and the model struggled with class imbalance, particularly in recognizing emotions like fear and sadness, where accuracy ranged between 58-60% [8].

This advanced research focuses on identifying human facial expressions in Indonesia using Convolutional Neural Networks (CNN) with the Indonesian Mixed Emotion Dataset (IMED). The IMED dataset consists of RGB images grouped into five categories, with 80% for training and 20% for testing. Data pre-processing includes face area cropping, image conversion to grayscale, and image dimension adjustment to 48x48 pixels. The network architecture consists of four-layer blocks, including a convolutional layer, activation layer (ReLU), normalization layer, pooling layer, and dropout layer. The initial stage uses a 3x3 kernel with 32 filters, resulting in a 48x48x32 feature map batch normalization and ReLU activation. Max-pooling with a 2x2 kernel is applied to reduce the spatial dimension. Subsequent blocks add filters to capture more complex information. In the final stage, the feature map is flattened into a 1D vector and processed through a fully connected layer for classification. This study reached a validation accuracy of 93.63% [9].

Research on the use of Convolutional Neural Networks (CNN) for facial expression recognition uses three datasets: FER2013, Cohn-Kanade (CK+), and Karolinska Directed Emotional Faces (KDEF). The CK+ dataset comprises 981 images, while KDEF has 490 images adjusted to 48x48 pixels. The model was developed with multiple convolution and fully connected layers and trained with various optimization scenarios and some epochs. As a result, the accuracy for the FER2013 dataset was 52% at 200 epochs and dropped to 49% at 500 epochs. For KDEF, the accuracy was 81% at 200 epochs and decreased to 77% at 500. The CK+ dataset showed 77% accuracy at 200 epochs and decreased to 71% at 500. When the three datasets were combined, the accuracy was 57% at 200 epochs and decreased to 54% at 500 epochs, indicating that variations in training and datasets affect the model performance [10]. This indicates limitations in the model's ability to generalize across different datasets and suggests potential overfitting, especially as performance declines with extended training.

A comparative study on FER with various machine learning techniques has been done to improve efficiency and accuracy. The methods tested include SVM, LR, ANN, RF, KNN, NB, and CNN, but the CNN architecture needs to be described. The evaluation was performed on ORL and Yale databases using accuracy, confusion matrix, and ROC performance measures. Results show that CNN and other deep learning models such as AlexNet, DenseNet, and LeNet achieve 100% performance on the ORL database, while traditional models such as SVM achieve 98.19%. However, deep learning techniques showed less satisfactory results on the smaller Yale dataset, emphasizing the importance of large datasets for optimal deep learning performance [11].

Research by Adrian et al. optimizes hyperparameters on CNN for facial emotion recognition using the FER2013 dataset, which is divided into training, validation, and test data with a ratio of 80-10-10. The proposed architecture includes five convolutional layers with 256, 512, 384, and 192 filter configurations and a dropout technique to reduce overfitting. Image pre-processing includes data augmentation. Results show significant improvement in accuracy, with a validation accuracy of 63.22% after 20 epochs and a test accuracy of 72.16% after 750 epochs. The model shows competitive performance with lower computational overhead than complex models such as VGG and ResNet; it is ranked ninth on the FER2013 benchmark on PapersWithCode [12]. Although the model achieved good performance on the test set, real-time application testing is not mentioned. It is unclear how well the model would perform when integrated into a real-time system, where speed and efficiency are crucial, especially on devices with limited computational resources.

Khaireddin et al. discussed facial emotion recognition using CNN architecture with a pre-trained VGGNet model on the FER2013 dataset divided by an 80:10:10 ratio for training, validation, and testing. The research emphasizes the effectiveness of CNN in automatic feature extraction and computational efficiency. The model uses four convolution layers and optimization techniques to improve performance and change the learning rate. The best accuracy achieved was 73.28% without additional training data, with Reduce Learning on Plateau callback optimization, and ranked eighth on the FER2013 accuracy benchmark in PapersWithCode [13].

Various previous studies show that Convolutional Neural Networks (CNN) technology has become the dominant method in facial expression recognition. Variations in the architecture, datasets, and training techniques used provide important insights into the effectiveness and adaptability of CNN models. These studies show that the depth and complexity of the CNN architecture, the number of neurons, and the selection of the right dataset greatly affect the model's performance, but also risk overfitting if not properly managed. There is great potential for

model development for facial expression classification, especially in the context of the Indonesian population, which may require models that are adaptive and relevant to local characteristics. In addition, to improve the model's generalizability, an effort will be made to combine the dataset of Western faces with non-Indonesian faces so that it is expected to produce a more diverse and adaptive model for the emotion category used.

III. METHODOLOGY

This research develops a Convolutional Neural Network (CNN)-based Facial Emotion Recognition (FER) model as shown in Fig. 1. In general, the process started with data collection and filtering of emotional data, followed by exploratory analysis to understand the characteristics of the data. Defective and duplicate images are removed to maintain the quality of the dataset. Images are then converted to grayscale, faces are detected using Haar Cascade and resized for uniformity. The processed data was divided into training and testing sets. The model was trained by adding a new convolution layer with a 1024 filter and using callbacks such as early stopping and ReduceLROnPlateau to optimize performance. Finally, the model was tested to evaluate its accuracy and performance in recognizing emotions from facial images.

A. Data Collection

This study uses secondary data from three primary datasets: CK+, FER2013, and IMED. The FER2013 dataset includes 35,882 facial images with a resolution of 48x48 pixels in grayscale format, grouped into seven emotion categories. There is no specific information about the gender or age of the subjects in the images. Fig. 2 displays each example image in each class of the FER2013 dataset [14] [15]. The figure shows images converted into grayscale format with a resolution of 48x48 pixels. When viewed, images in the FER2013 dataset are covered by watermarks and hand poses that cover the face, which is expected to interfere with the learning process of the model and affect the accuracy. In addition, some emotion classes may need to be better balanced, with some emotions, such as disgust, having a much smaller number of samples than other emotions, such as happy or sad.

The CK+ dataset consists of 981 facial expressions from 210 individuals aged 18-50, with a gender distribution of 31% male and 69% female and a racial distribution of 81% Euro-American, 13% Afro-American, and 6% other. The dataset has seven emotion categories, and the images are in grayscale format with 48x48 pixel resolution. Fig. 3 shows each example image in each class of the CK+ dataset [10]. The number of images in each emotion class was observed, showing that this dataset has the least number compared to the other datasets used in this study. The pixel size of each image is 48x48 pixels and is in grayscale. The facial expressions of each emotion category are very clearly identified. In addition, the images in this dataset were taken under good and controlled lighting conditions, and all subjects were oriented directly toward the camera, which facilitated the analysis of facial expressions.

The IMED dataset contains 9,183 images of six male and nine female Indonesian subjects aged 17-32 from various Java, Batak, Sunda, Minang, and Manado ethnicities. The dataset comprises seven emotion categories and is organized in grayscale format with a resolution of 720x480 pixels. This dataset was obtained with special permission through the website https://imed.cs.ui.ac.id/. Fig. 4 shows each example image in each class of the IMED dataset [16]. It can be seen that the images in the IMED dataset have a resolution of 720 pixels wide and 480 pixels high, with an RGB image mode indicating full-color representation. The image quality in the IMED dataset has good and uniform lighting, and all the datasets are clean, with no noise, and are not covered by watermarks or hand poses covering the face area. However, the images are framed in black frames, indicating that they are still raw. Therefore, it is necessary to pre-process the data to produce uniform images with the other datasets, which ideally have a size of 48x48 pixels and are in grayscale mode.

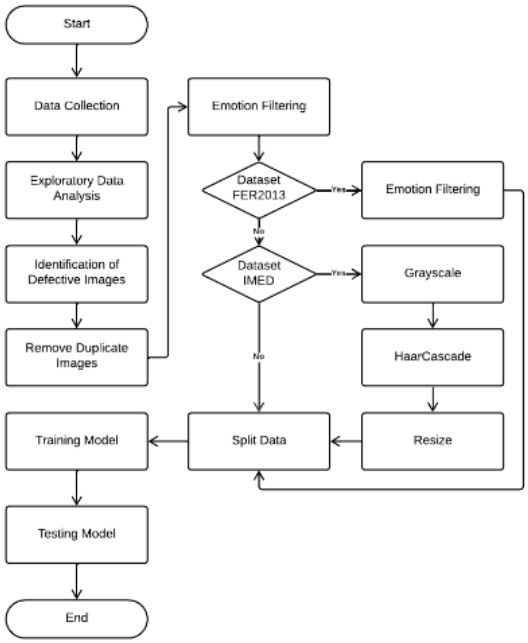


Fig. 1. Face recognition flowchart.

Fig. 2. FER2013 dataset images example.

Fig. 3. CK+ dataset images example.

Fig. 4. IMED dataset images example.

Table I shows the calculation details of each emotion class in the dataset.

TABLE I. CLASS DISTRIBUTION IN EACH DATASET

Emotion Class	FER2013	IMED	CK+
Neutral	6.193	518	54
Angry	4.953	1.623	135
Disgust	5.470	1.413	177
Fear	5.121	1.466	75
Happy	8.989	1.319	207
Sad	6.077	1.793	84
Surprise	4.002	1.051	249
Total	35.882	9.183	981

B. Data Preparation

In this study, the data pre-processing steps were carried out with a special focus on the IMED dataset, which is still in raw form, unlike the other datasets, which have been converted to grayscale and adjusted to 48x48 pixel dimensions. Details on the processing of the IMED dataset are shown in Fig. 5.

- Face detection and cropping of the image using the Haar Cascade Classifier method from the OpenCV library. Face detection is performed to ensure that only parts of the face are analyzed. This method compares parts of the image with pre-trained facial features like eyes, nose, or mouth. Once the face is detected, the area is cropped for face isolation [17]
- Grayscale conversion. In the IMED dataset, the image is converted to grayscale mode. This conversion reduces the complexity of the data but still retains the important features of the face. This process uses the cv2.cvtColor function with the cv2.COLOR_BGR2GRAY parameter.
- Resize Image: after the image is converted to a grayscale format, the image with variable sizes is resized to 48x48 pixels to maintain consistency with other datasets and ease of model processing.

Fig. 5. IMED dataset pre-processing steps.

In the FER2013 dataset, the data provided in CSV format requires conversion from a string of pixel values to an array of images. This process involved using np. from string to convert the string into a one-dimensional integer array, which was then reshaped into an image with a resolution of 48x48 pixels. This reshaping technique converts the data into a pixel matrix that the CNN model can process. A visualization of the processing of the FER2013 dataset is shown in Fig. 6.

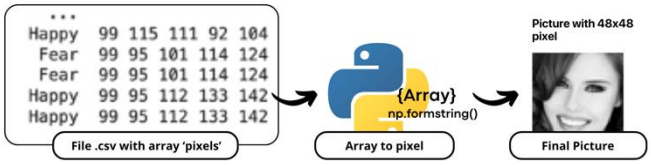


Fig. 6. FER2013 dataset pre-processing steps.

Afterward, emotion selection was conducted to simplify the emotion categories from 7 to 4. The four emotions are happy, angry, neutral, and sad. After emotion selection, the data was divided with a ratio of 80:18:2 for training, validation, and test data. The test data was then expanded by adding images taken from photos of real people, which was done to ensure a more genuine representation. The downsampling process balances the amount of test data between each emotion label. Two experiments were conducted with dataset splitting. The first experiment is conducted with the number of emotions in each dataset intact, with no reductions. Table II shows the training, validation, and test data used in the first experiment.

TABLE II. NUMBER OF TRAINING, VALIDATION, AND TESTING DATA

Data Split	Angry	Happy	Neutral
Training	5.160	8.208	5.240
Validation	1.132	1.856	1.183
Testing	148	148	148
Total	6.440	10.212	6.571

C. Model Architecture

This research uses the Convolutional Neural Network architecture. The CNN architecture used is adapted with the addition of one layer of convolutional blocks based on references from the PapersWithCode website, specifically from research entitled "Convolutional Neural Network Hyperparameters Optimization for Facial Emotion Recognition" [12], with details observable in Fig. 7.

The model starts with an input layer, where a grayscale and normalized face image of 48x48 pixels is used as input. Next, a series of convolution blocks extract visual features from the image. Each convolution block consists of a 2D Convolution layer with 3x3 filters; the first block has 256 filters, the second block 512, the third block 384, the fourth block 192, and the fifth block 512. The output of each convolution layer is normalized using BatchNormalization, and its dimension is reduced through MaxPooling2D with a 2x2 window. A dropout of 0.5 is applied after pooling to prevent overfitting.

Fully Connected Layers in this model include several layers: a Flatten Layer, which converts the multidimensional output of the last convolutional block into a one-dimensional vector, followed by a Dense Layer with 256 neurons and ReLU activation. A dropout of 0.5 is applied to the Dense Layer output to prevent overfitting. Finally, the Output Dense Layer uses four neurons corresponding to the number of emotion classes, equipped with a softmax activation function to generate class probabilities.

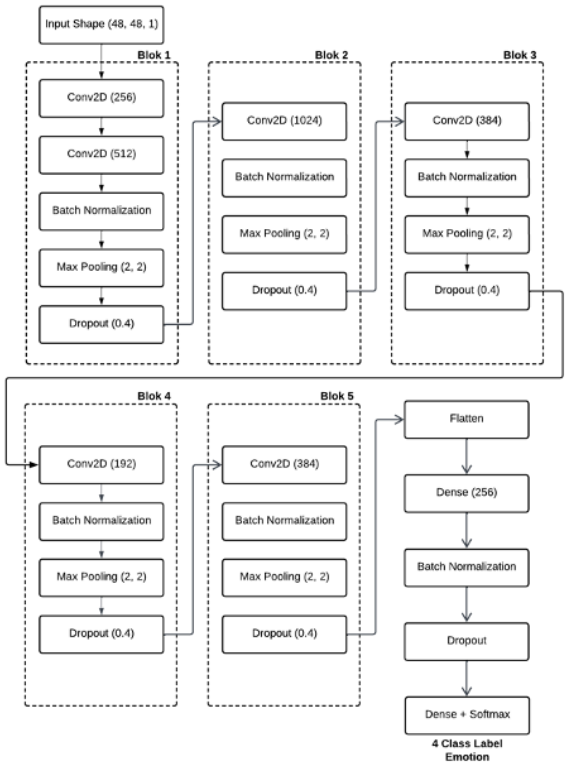


Fig. 7. Implemented CNN architecture.

D. Testing Scenario

Before going into the training and testing phase of the model, nine test scenarios were conducted to evaluate the model's performance under various conditions. These scenarios were designed to understand how datasets and model architecture variations affect the accuracy and loss of training and validation data. Table III summarizes the nine test scenarios conducted.

TABLE III. TESTING SCENARIO

Scenario	Description
1	Train using IMED dataset only
2	Train using CK+ dataset only.
3	Train using the FER2013 dataset only.
4	Train using combined IMED and FER2013 datasets
5	Train using combined IMED and CK+ datasets
6	Train using combined FER2013 and CK+ datasets
7	The first train without additional convolution layers with no minimum count using the FER2013 dataset
8	The second train without additional convolution layers with 1000 per label using the FER2013 dataset
9	The third train with the addition of a convolution block with a 1024 filter using the FER2023 dataset

Every scenario is designed to explore different aspects of model training. Scenarios 1 to 3 aim to measure the model's performance trained with individual datasets such as IMED, CK+, and FER2013. Scenarios 4 to 6 evaluate combinations of datasets to see the impact of integrating data from different sources. Scenarios 7 and 8 investigate the effects of variations in the amount of data and the use of additional convolution layers on the FER2013 dataset. Finally, scenario 9 involves adding convolution layers with large filters, batch normalization, max pooling, and dropout to capture more complex features and reduce overfitting. These scenarios helped identify the best configuration to implement the facial emotion recognition model.

E. Callback Mechanism

A callback method is used to optimize the process in model training using Keras TensorFlow. An early stopping callback is applied to automatically stop training if there is no improvement in validation accuracy [18] [19]. This allows the model to stop by seven earlier than the maximum number of epochs if the model is no longer improving, thus saving computational time and resources. Model saving is done through Model Checkpoint, which only saves the model if there is an increase in validation accuracy; this ensures that the saved file is optimal based on validation accuracy. Furthermore, the ReduceLROnPlateau callback method reduces the learning rate of the model by 80% from its previous value if there is no improvement in validation accuracy for three consecutive epochs, considering the minimal change in validation accuracy of 0.0001 to be regarded as an improvement [20]. Changes smaller than this value will not be considered.

The model is compiled with a loss function or sparse categorical cross-entropy loss function suitable for classification tasks where the target class label is an integer [21][22]. The optimizer used is Adam, which is used with early stopping callbacks, and ReduceLROnPlateau to optimize the training process with an initial learning rate of 0.001 [23]. The metric measured is accuracy, which indicates the percentage of correct predictions. Training is performed with the training dataset and validation data to evaluate model performance on data not used during training. During training, the model will go through a maximum of 50 epochs, which implies that the data will be processed 50 times to optimize the model weights based on the loss function and the accuracy measurement.

IV. RESULT AND DISCUSSION

After performing duplicate and corrupted image detection on each dataset, 3369 duplicate images were found on the FER2013 dataset and 273 duplicate images on the IMED dataset. No corrupted images were found in any of the analyzed datasets. In the image pre-processing stage, nine images from the IMED dataset are not detected by the Haar Cascade Classifier, as shown in Fig. 8. The duplicate images that the Haar Cascade Classifier does not detect will be removed to ensure the consistency and quality of the data used in training the model.

Fig. 8. Example of images that Haar Cascade Classifier does not detect.

After the pre-processing stage is complete, the next step is to perform model training using a Convolutional Neural Network (CNN). This training is done to optimize the model in recognizing facial emotions from images. Several experiments according to the design of the test scenario were conducted to evaluate the model's performance on various combinations of datasets, both individually and combined, and to explore the effect of the model architecture on its performance.

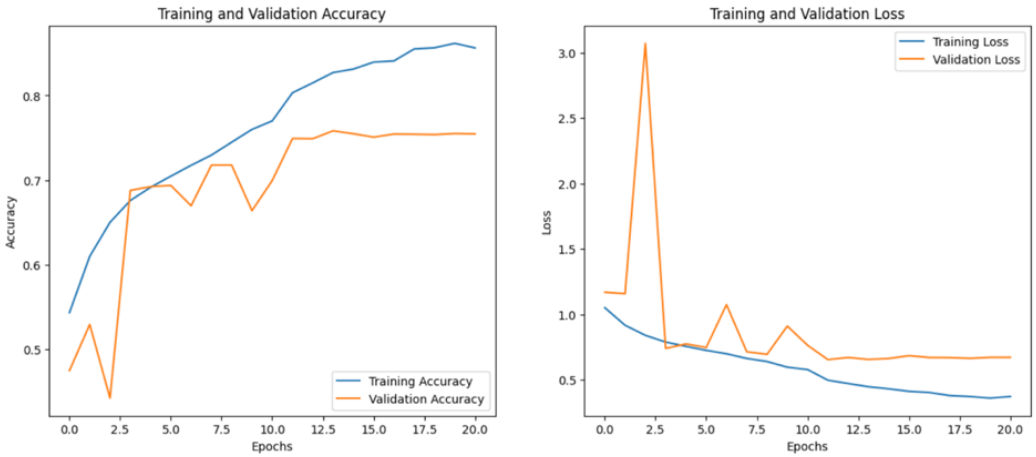


Fig. 9. Scenario 7: Train-validation accuracy and loss plot (training process was stopped early at the 21st epoch).

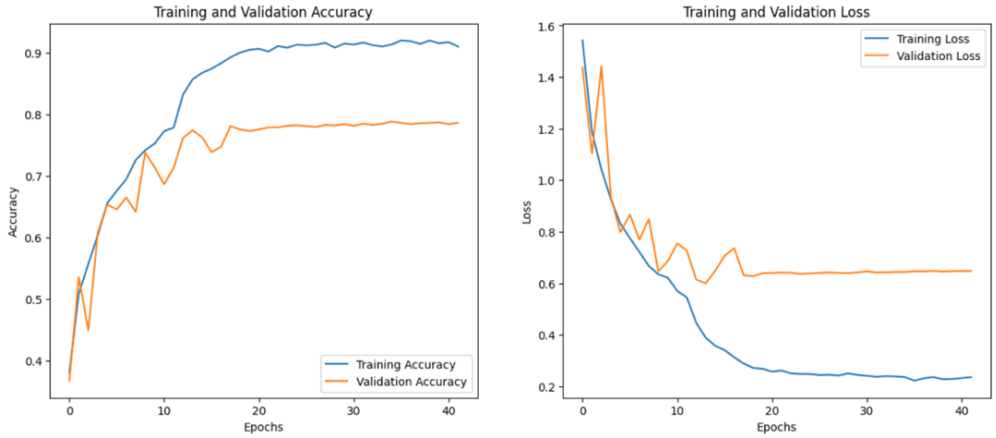


Fig. 10. Scenario 8: Train-validation accuracy and loss plot (each epoch took an average of 24 seconds).

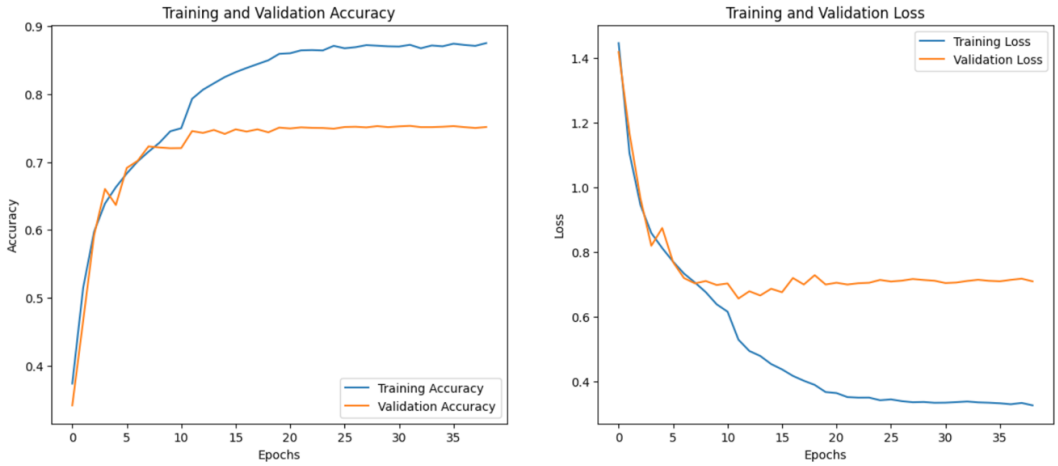


Fig. 11. Scenario 9: Train-validation accuracy and loss plot (validation loss shows a more stable decrease).

In the first experiment of training the combined model of three datasets, the highest validation accuracy was obtained at the 14th epoch with an accuracy value of 0.7582. At the 14th epoch, the model also recorded a validation loss of 0.6572, one

of the lowest loss values during the training session. For each epoch, the training took 49 seconds. The lowest validation loss value was recorded at the 12th epoch with a value of 0.6556, while the highest was recorded at the 3rd epoch with a very high value of 3.0713. This shows significant fluctuations in the model's performance at the beginning of training. During the training session, the model decreased the learning rate three times, precisely at the 11th, 17th, and 20th epochs. This decrease is done automatically by the ReduceLROnPlateau callback, which aims to improve validation accuracy when stagnation in performance improvement is detected. In addition, the more epochs performed, the higher the validation loss. The training process was stopped early at the 21st epoch, as shown in Fig. 9.

The training process of the second combined dataset lasted for 42 epochs before being stopped by the early stopping mechanism. Each epoch had 227 batches of data used for training. Furthermore, the early stopping method was set to stop training if there was no improvement in validation accuracy for 7 consecutive epochs, restoring the model weights to the state when the best performance was achieved. The model training started with a high initial loss of 15.42%, an accuracy of 38.03% on the training data, and about 36.69% on the validation data. Significant improvement occurred from the start of training, with training accuracy increasing to 75.24% and validation accuracy reaching 73.76% by the 9th epoch. The biggest improvement occurred at the 24th epoch, where the validation accuracy soared from 66.47% in the previous epoch to 78.07%. Each epoch took an average of 24 seconds. The accuracy results are shown in Fig. 10.

Furthermore, for the final experiment, adding a convolution layer with a filter of 1024, batch normalization, max pooling, and dropout resulted in a slight increase in model accuracy on the validation data. With this new configuration, the model achieved a validation accuracy of 79.59% after 39 epochs, with early stopping at the 39th epoch. This is compared to previous experiments that showed an increase in accuracy from 75.82% to 78.81% after data and hyperparameter adjustments. Adding a new convolutional layer with a large filter gives the model additional capacity to capture more complex features, thus improving classification ability. In addition, the validation loss of the new model shows a more stable decrease, as shown in Fig. 11.

TABLE IV. TRAINING RESULTS OF ALL SCENARIOS

Scenario	Train Accuracy	Train Loss	Validation Accuracy	Validation Loss
1	0.9855	0.1297	0.9809	0.0877
2	0.9766	0.0800	0.9792	0.1037
3	0.8272	0.4393	0.7226	0.7922
4	0.8637	0.3551	0.7575	0.7100
5	0.9716	0.0792	0.9859	0.1060
6	0.3018	0.8822	0.7178	0.8671
7	0.8271	0.4488	0.7582	0.6572
8	0.9135	0.2367	0.7881	0.6440
9	0.9339	0.1814	0.79558	0.6206

Table IV summarizes the model's best training results on four different datasets, namely IMED, CK+, FER2013, and a combined dataset, showing the variation in model performance in terms of accuracy and loss for both training and validation data at various points of the last epoch when training is stopped. This study shows that the model's accuracy decreases when the FER2013 dataset is mixed with other datasets, such as CK+ or IMED. The combination of FER2013 and CK+ resulted in a lower validation accuracy of 71.78%, with a high validation loss of 86.71% at the 39th epoch. Although the CK+ dataset is of high quality, the presence of FER2013 with high variability reduces the model's overall performance. The combination of IMED and FER2013 showed a validation accuracy of 75.75% and a validation loss of 71%. Although there is a performance improvement compared to using FER2013 alone, the results are still lower than those of using the IMED dataset individually. The CK+ and IMED datasets are more homogeneous and controlled, with more consistent expression variation and higher image quality. When the more varied and uncontrolled FER2013 dataset is mixed with the more homogeneous dataset, the model needs help generalizing the relevant patterns, thus lowering the overall performance. In the first and second experiments, no additional convolution layer was used.

Considering the results in Table IV for the three combined datasets in scenarios 7, 8, and 9, the best model performance is achieved in scenario 9. Therefore, the final model selected for further analysis is scenario 9. A confusion matrix study is conducted to better understand the model's performance in facial emotion classification. The confusion matrix provides an overview of how well the model classifies each emotion category and identifies areas that require further improvement [5] [24]. This analysis is important to evaluate the accuracy and weaknesses of the model in emotion recognition and to find out the most frequent types of errors.

TABLE V. TESTING RESULT CONFUSION MATRIX

	Predicted				
	Class	**Angry**	**Happy**	**Neutral**	**Sad**
True	Angry	32	3	10	2
	Happy	0	41	3	3
	Neutral	6	5	32	4
	Sad	9	3	7	28

The confusion matrix in Table V shows the model's performance in classifying facial emotions into four categories: angry, happy, sad, and neutral. For the original label angry, the model successfully classified 32 samples correctly, but ten samples were misclassified as neutral, three as happy, and two as sad. The happy original label performed better, with 41 samples correctly classified, although three samples were misclassified as neutral and three as sad. For the neutral label, 32 samples were correctly classified, while six were incorrectly classified as angry, five as happy, and four as Sad. Finally, the model classified 28 samples correctly for the original label sad, but there were errors with nine samples classified as angry, three as happy, and seven as neutral. This analysis shows that the model is quite accurate in classifying happy and neutral emotions but still needs improvement in classifying angry and

sad feelings, especially in distinguishing between angry and neutral.

TABLE VI. CLASS CLASSIFICATION RESULT

	Precision	Recall	F1-Score	Support
Angry	0.68	0.68	0.68	47
Happy	0.79	0.87	0.83	47
Neutral	0.65	0.68	0.65	47
Sad	0.76	0.60	0.67	47
Accuracy			0.71	188
Macro average	0.71	0.71	0.71	188
Weighted average	0.71	0.71	0.71	188

The results of the emotion classification metrics in Table VI show the model's precision, recall, and F1-Score performance for the angry, happy, sad, and neutral categories. For anger, the model achieved a precision and recall of 0.68. The happy category shows the best performance by the highest precision of 0.79, recall of 0.87, and F1-score of 0.83. The neutral category gets a precision of 0.65 and a recall of 0.68, while the sad category gets a precision of 0.76 and a recall of 0.60 with an F1-Score of 0.67. The model's overall accuracy was 0.71, with a macro-average and weighted average of 0.71 each.

Furthermore, the model is analyzed using the ROC Curve. The ROC curve illustrates the performance of the classification model at all possible classification thresholds by plotting two parameters: True Positive Rate (TPR) and False Positive Rate (FPR). AUC is the probability that the model will rank random positive samples higher than random negative samples. The ROC curve AUC is presented in the form of a curve graph. AUC values range from 0 to 1, where 1 indicates perfect prediction, and 0.5 indicates performance no better than a random guess [25].

Fig. 12 shows the ROC Curve in the range of 0-1 (higher is better) for scenario 9. The X axis is the false positive rate in the range of 0-1 and the Y axis is the true positive rate in the range of 0-1. The ROC shows the best performance for the happy class (AUC 0.97) and good for the angry and sad class (AUC 0.90). The neutral class has the lowest AUC of 0.87.

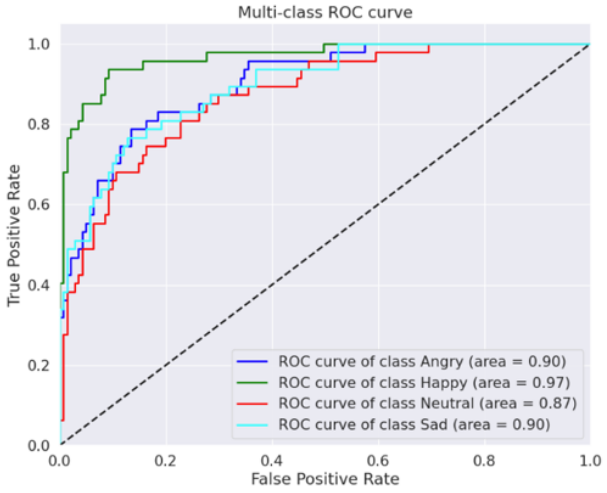
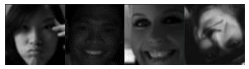
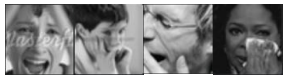


Fig. 12. ROC Curve on Scenario 9 classes.

The dataset characteristics affect the model's effectiveness, with variability in lighting and image quality making it difficult to extract important features. Table VII shows that images with unclear expressions or partially obscured faces degrade the model's performance, and poorly lit or non-face images hinder accurate classification.

TABLE VII. NON-CONFORMING IMAGES

Remarks	Images
Images with cartoon face	
Poorly lit images	
Face area covered by hands or watermarks	
Non-face images	

V. CONCLUSION

This research successfully developed a Facial Emotion Recognition model using Convolutional Neural Network (CNN) architecture by combining CK+, FER2013, and IMED datasets. In the 9th scenario, the model incorporating the three datasets with an additional convolutional layer of 1024 filters achieved the highest validation accuracy of 79.59%. The main factors for improved accuracy are better feature extraction capabilities and the application of batch normalization, max pooling, and dropout, which reduce overfitting and improve validation stability. The model achieved an average accuracy, precision, recall, and F1-Score of 71.5% on the test data.

Using the IMED and CK+ datasets individually provides high validation accuracy (97-98%), while the more varied FER2013 dataset provides a validation accuracy of 72.64%. Training parameter adjustments such as learning rate reduction and early stopping also improved model performance. This research shows that combining datasets and improved CNN architecture significantly affects the performance of facial emotion recognition.

ACKNOWLEDGMENT

We would like to thank: Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Kristen Duta Wacana, Indonesia for the grant support No. 108/D.01/LPPM/2024.

REFERENCES

[1] T. Kumar Arora et al., "Optimal Facial Feature Based Emotional Recognition Using Deep Learning Algorithm," Comput Intell Neurosci, vol. 2022, 2022, doi: 10.1155/2022/8379202.

[2] S. Zhao, Y. Gao, X. Jiang, H. Yao, T. S. Chua, and X. Sun, "Exploring principles-of-art features for image emotion recognition," in MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia, Association for Computing Machinery, Nov. 2014, pp. 47–56. doi: 10.1145/2647868.2654930.

[3] Purwono, A. Ma'arif, W. Rahmaniar, H. I. K. Fathurrahman, A. Z. K. Frisky, and Q. M. U. Haq, "Understanding of Convolutional Neural Network (CNN): A Review," International Journal of Robotics and

Control Systems, vol. 2, no. 4, pp. 739–748, 2022, doi: 10.31763/ijrcs.v2i4.888.

[4] L. Bejjagam and R. Chakradhara, "Facial Emotion Recognition using Convolutional Neural Network with Multiclass Classification and Bayesian Optimization for Hyper Parameter Tuning," 2022. [Online]. Available: www.bth.se

[5] B. C. Ko, "A brief review of facial emotion recognition based on visual information," Sensors (Switzerland), vol. 18, no. 2, Feb. 2018, doi: 10.3390/s18020401.

[6] J. U. Rahman, F. Makhdoom, and D. Lu, "ASU-CNN: An Efficient Deep Architecture for Image Classification and Feature Visualizations."

[7] L. Alzubaidi et al., "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," J Big Data, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00444-8.

[8] K. Liu, M. Zhang, and Z. Pan, "Facial Expression Recognition with CNN Ensemble," in Proceedings - 2016 International Conference on Cyberworlds, CW 2016, Institute of Electrical and Electronics Engineers Inc., Nov. 2016, pp. 163–166. doi: 10.1109/CW.2016.34.

[9] A. Bayu Jala, "Implementasi Algoritma Convolutional Neural Network (CNN) Untuk Klasifikasi Ekspresi Wajah Manusia Di Indonesia," Bandung, 2020. Accessed: Oct. 09, 2023. [Online]. Available: https://repository.telkomuniversity.ac.id/pustaka/165458/implementasi-algoritma-convolutional-neural-network-cnn-untuk-klasifikasi-ekspresi-wajah-manusia-di-indonesia.html

[10] F. Azizi Nur, "Deteksi Emosi Menggunakan Citra Ekspresi Wajah Secara Otomati," 2021.

[11] A. A. Komlavi, K. Chaibou, and H. Naroua, "Comparative study of machine learning algorithms for face recognition", doi: 10.46298/arima.9291ï.

[12] A. Vulpe-Grigorasi and O. Grigore, "Convolutional Neural Network Hyperparameters optimization for Facial Emotion Recognition," in 12th International Symposium on Advanced Topics in Electrical Engineering, ATEE 2021, Institute of Electrical and Electronics Engineers Inc., Mar. 2021. doi: 10.1109/ATEE52255.2021.9425073.

[13] Y. Khaireddin and Z. Chen, "Facial Emotion Recognition: State of the Art Performance on FER2013."

[14] L. Pham, T. H. Vu, and T. A. Tran, "Facial expression recognition using residual masking network," in Proceedings - International Conference on Pattern Recognition, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 4513–4519. doi: 10.1109/ICPR48806.2021.9411919.

[15] L. Zahara, P. Musa, E. Prasetyo Wibowo, I. Karim, and S. Bahri Musa, "The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi," in 2020 5th International Conference on Informatics and Computing, ICIC 2020, Institute of Electrical and Electronics Engineers Inc., Nov. 2020. doi: 10.1109/ICIC50835.2020.9288560.

[16] D. Y. Liliana, T. Basaruddin, and I. I. D. Oriza, "The Indonesian Mixed Emotion Dataset (IMED): A facial expression dataset for mixed emotion recognition," in ACM International Conference Proceeding Series, Association for Computing Machinery, Nov. 2018, pp. 56–60. doi: 10.1145/3293663.3293671.

[17] S. Yulina, "100-109 Dokumen diterima pada 21 Januari," 2021. [Online]. Available: https://jurnal.pcr.ac.id/index.php/jkt/

[18] F. Fischer, A. Birk, P. Somers, K. Frenner, C. Tarín, and A. Herkommer, "FeaSel-Net: A Recursive Feature Selection Callback in Neural Networks," Mach Learn Knowl Extr, vol. 4, no. 4, pp. 968–993, Dec. 2022, doi: 10.3390/make4040049.

[19] S. Samidin and A. Fadjeri, "Klasifikasi Gambar Batu-Kertas-Gunting Menggunakan Convolutional Neural Network dengan Fungsi Callback untuk Mencegah Overfitting," Jurnal Penelitian Inovatif, vol. 4, no. 2, pp. 785–794, Jun. 2024, doi: 10.54082/jupin.413.

[20] Y. Wu and L. Liu, "Selecting and Composing Learning Rate Policies for Deep Neural Networks," ACM Trans Intell Syst Technol, vol. 14, no. 2, Feb. 2023, doi: 10.1145/3570508.

[21] S. Chatterjee and A. Keprate, "Predicting Remaining Fatigue Life of Topside Piping Using Deep Learning," in 2021 International Conference on Applied Artificial Intelligence, ICAPAI 2021, Institute of Electrical and Electronics Engineers Inc., May 2021. doi: 10.1109/ICAPAI49758.2021.9462055.

[22] B. N. Chaithanya, T. J. Swasthika Jain, A. Usha Ruby, and A. Parveen, "An approach to categorize chest X-ray images using sparse categorical cross entropy," Indonesian Journal of Electrical Engineering and Computer Science, vol. 24, no. 3, pp. 1700–1710, Dec. 2021, doi: 10.11591/ijeecs.v24.i3.pp1700-1710.

[23] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Dec. 2014, [Online]. Available: http://arxiv.org/abs/1412.6980

[24] Z. Karimi, "Confusion Matrix," 2021. [Online]. Available: https://www.researchgate.net/publication/355096788

[25] M. Tripathi, "FACIAL EMOTION RECOGNITION USING CONVOLUTIONAL NEURAL NETWORK," ICTACT JOURNAL ON IMAGE AND VIDEO PROCESSING, p. 1, 2021, doi: 10.21917/ijivp.2021.0359.

Optimization of 3D Coverage Layout for Multi-UAV Collaborative Lighting in Emergency Rescue Operations

Dan Jiang[1], Rui Yan[2]*
Beijing Municipal Commission of Housing and Urban-Rural Development Beijing, China[1]
School of Economics and Management, University of Science and Technology Beijing, Beijing, China[2]

***Abstract*—In emergency rescue scenarios, Unmanned Aerial Vehicles (UAVs) play a pivotal role in navigating complex terrains and high-risk environments. This paper proposes an optimization model for the three-dimensional coverage layout of a multi-UAV collaborative lighting system, specifically designed to meet the spatial requirements of emergency operations. An enhanced Particle Swarm Optimization (PSO) algorithm is employed to tackle the layout challenges, featuring adaptive inertia weights and asymmetric learning factors to improve both efficiency and global search capabilities. The simulation results demonstrate that the proposed method significantly enhances coverage efficiency, achieving over 90% coverage in critical areas while ensuring precise UAV positioning. Additionally, the algorithm shows faster convergence and stronger global search ability, effectively optimizing UAV deployment and improving operational efficiency during rescue missions. This study offers a practical and reliable layout solution for multi-UAV collaborative lighting systems, which is crucial for reducing rescue times, ensuring operational safety, and improving resource allocation in emergency responses.**

Keywords*—Unmanned Aerial Vehicles; emergency rescue; collaborative lighting; three-dimensional coverage; particle swarm algorithm

I. Introduction

Geological disasters are increasingly prevalent worldwide, with the Global Disaster Data Platform reporting 392 wildfires, 109 earthquakes, and 109 floods in the last six months alone [1]. These statistics likely underestimate the true frequency of such events. The sudden and unpredictable nature of these disasters, coupled with complex environmental conditions, poses significant challenges for emergency rescue operations. Issues such as injuries to personnel, transportation difficulties, and disruptions in communication, power, and lighting infrastructure complicate these efforts further.

In this context, UAVs have emerged as crucial tools in managing such crises. UAVs, which do not require an onboard human presence, offer flexibility, efficiency, and convenience, proving especially useful in confined spaces, complex environments, or hazardous conditions. Their deployment can significantly reduce rescue times, enhance operational efficiency, and improve post-disaster relief efforts [2]. UAVs have broad applications across various sectors such as agriculture, construction, transportation, emergency rescue, military operations, and collaborative lighting [3].

The use of UAVs is particularly critical in emergency rescue situations where traditional lighting equipment may fail to provide rapid and effective illumination [4]. A collaborative approach using multiple UAVs can achieve more flexible and efficient lighting solutions. This paper focuses on optimizing the three-dimensional coverage layout of UAVs equipped with lighting systems, which play pivotal roles in emergency scenarios by enhancing lighting conditions, expanding coverage areas, improving safety, easing workload, and offering flexibility to adapt to diverse mission requirements.

The contributions of this paper include enhancing the comprehensiveness and inclusiveness of the 3D coverage model through innovative UAV technology that depicts disaster scenes and obstacle locations, optimizing 3D lighting coverage using an advanced particle swarm optimization algorithm to improve calculation accuracy and efficiency, and providing a comprehensive discussion on the role of UAV lighting systems in rescue operations. This optimization aids in reducing rescue times, ensuring the safety of rescuers, and minimizing the consumption of resources.

The structure of the paper is organized as follows: Section II reviews related literature; Section III develops the 3D coverage model for the lighting system; Section IV introduces the enhanced particle swarm optimization algorithm; Section V presents a computational study with examples and numerical results; Section VI concludes the paper.

II. Literature Review

A. Current Development of UAVs and their Applications in Different Scenarios

Over the past few decades, UAVs have rapidly evolved, transitioning from their initial military use to various civilian applications. One of the research areas receiving significant attention from scholars is the application of UAVs in emergency rescue scenarios. Different types of emergency rescue UAVs can be utilized for ground site exploration, thermal imaging search and rescue, rescue site mapping, and manned evacuations. In emergency rescue operations, UAVs offer valuable support for information gathering and complex environment rescue due to their compact size, flexibility, and user-friendliness [5]. Moreover, the modular design of UAVs enables the attachment of different systems tailored to specific emergency scenarios, increasing the range of rescue missions.

UAV applications are rapidly expanding across multiple domains. In the realm of smart cities and public safety, UAVs serve as versatile mobile platforms for smart city applications and security surveillance. The utilisation of UAVs and vehicles for last-mile urban deliveries represents an innovative delivery approach. This delivery model's optimisation problem is recognized as the UAV Travelling Salesman Problem (TSP-D). El-Adle et al. [6] studied a deterministic last-mile delivery problem by designing heuristic algorithms based on mixed integer programming to efficiently and accurately prioritise customers for UAV deliveries. Li et al. [7] introduced trucks and a UAV routing model (TDRP-FT) that considers flight altitude selection and time windows. The TDRP-FT model is validated using a hybrid algorithm that combines a greedy stochastic adaptive search process and an adaptive large-neighbourhood search to assess its effectiveness [8].

UAVs can significantly enhance urban public safety by offering early warning and support in diverse emergency scenarios and security operations. Based on BeiDou satellite positioning technology and Lora communication technology, [9] proposed a method of using multiple single-target trackers to replace part of the yolov3 detection task,[10] and put forward the design idea and specific implementation scheme of the vehicle monitoring prediction model.

Urban UAV route planning is a pivotal challenge, involving the quest for an optimal flight path under specific conditions. Miao et al. [11] proposed an improved symbiotic organism search algorithm based on the Simple Shape Method (SMSOS) to solve the path planning problem for UAVs, which has faster convergence speed, higher accuracy, and stronger robustness. Cui et al. [12] proposed a UAV trajectory planning scheme based on the hp-adaptive Radau pseudo-spectral method, considering both the dynamics constraints of the UAV and the terrain obstacle constraints during flight, and demonstrated its efficacy. Zhou et al. [13] proposed a mixed-integer linear programming model to address the reinitialized resource-constrained shortest path problem for correcting UAV flight errors. Three pruning strategies are introduced to enhance the effectiveness and efficiency of the algorithm, mitigating the high computational cost associated with the impulse algorithm. Shima and Schumacher [14] adopt a genetic algorithm that efficiently searches the feasible solution space to optimise the combination of multiple UAVs executing collaborative tasks, effectively solving the coupled problem of UAV task assignment and path planning.

UAVs have consistently played a significant role in the realm of civil security and disaster response. Most UAV-related research has made significant advancements in the domains of earthquakes, volcanic eruptions, and landslides. Following an earthquake, UAVs can serve as valuable investigative tools for collecting data from buildings that are inaccessible or require inspection. Li et al. [15] introduced an unsupervised method for detecting roof damage using UAV-captured images, which leverage combined color and shape features to assess the earthquake's impact based on the images acquired by UAVs. Thiels et al. [16] investigated the application of UAVs in the transportation of emergency supplies to disaster-stricken areas, hospitals, and offshore vessels during critical situations. Nagasawa et al. [17] proposed a multi-UAV coverage path planning method for 3D reconstruction of damaged buildings after a disaster. By generating, filtering, and ordering camera positioning points around a target damaged building, a final path that balances flight distance and time is provided for each UAV. Guo et al. [18] investigate optimal path planning for UAVs navigating in 3D environments with obstacles. They frame the problem as a nonlinear optimal control problem, incorporating continuous state inequality and terminal equivalence constraints. They propose a control parameterisation-based method to solve it. Ma et al. [19] design specific scenarios to simulate real disaster relief tasks and investigate UAV disaster relief task assignment problems. They propose three algorithms based on the horseshoe crab reagent algorithm: Aggregation-LAL, which utilizes an aggregation mechanism; Dynamic Adjustment-LAL, which employs dynamic adjustment mechanism; and Aggregation and Adjustment algorithms. Results demonstrate that AA-LAL achieves an average 5% improvement in task allocation in special scenarios compared to existing algorithms.

B. Coverage Metrics of the Lighting System in 3D Space

When considering three-dimensional lighting system coverage, the initial consideration is the selection of the light source to enable the measurement of light intensity and range. This paper utilizes LED light sources for illumination in a multi-UAV-carrying lighting system. LED light sources is known for their extended lifespan, compact size, and high luminous efficiency. Research on achieving uniform LED light intensity primarily involves two aspects: the secondary design of the LED light source and optimizing the arrangement and spacing of the LEDs.

The two-dimensional coverage problem has been a prominent subject of prior research in sensor networks. While most studies employ 2D models for network coverage, it is increasingly recognized that these models do not align well with real-world scenarios. Consequently, research on the 3D spatial coverage problem is gaining momentum. Huang et al. [20] (2007) were the first to carry out the research on 3D coverage, and they calculated the coverage by discussing the k-coverage problem. Alam and Haas [21] addressed the problem of 3D coverage and connectivity in underwater networks by dividing the 3D network volume into virtual cells in a distributed and scalable manner and selecting dynamic nodes to achieve perceptual coverage. The scheme can be used to achieve k-coverage. Gou et al. [22] proposed an energy efficient coverage enhancement method VKECE - 3D based on 3D - Voronoi partitioning and K - means algorithm, which guarantees the coverage while keeping the number of active nodes to a minimum. VKECE - 3D improves the network coverage and prolongs the network lifetime considerably. Li et al. [23] introduced a novel architecture for wireless sensor networks that incorporated distributed algorithms to manage sensor depths using an adaptive algorithm, resulting in maximal 3D sensor space coverage. Ramesh [24] presented a 3D spatial path planning technique for UAV-based atmospheric environment detection. This method can generate paths with complete coverage and optimal coverage density while accounting for energy constraints. Furthermore, it offers enhanced coverage density when compared to other path generation methods.

Currently, addressing the three-dimensional coverage problem involves mainly either converting the three-dimensional space into a two-dimensional space for computation or developing algorithms specifically designed for three-dimensional space coverage. Nevertheless, achieving accurate coverage calculations in the context of 3D coverage models necessitates more comprehensive and in-depth research efforts.

C. Optimisation Algorithm for the Objective Function

The optimisation of objective functions is a prevalent research area. In recent decades, researchers have developed a range of optimisation algorithms to tackle diverse challenges and problem types.

Classical single-objective optimisation algorithms encompass gradient descent, genetic algorithms, ant colony optimisation, and particle swarm optimisation, among others. Kennedy and Eberhart [25] introduced Particle Swarm Optimisation (PSO), an optimisation algorithm rooted in group intelligence. PSO achieves the search for a globally optimal solution by simulating the collaborative behaviour observed in natural phenomena like bird flocks or fish schools. The application of the particle swarm optimisation algorithm in objective optimisation problems represents an intriguing research avenue.

Nevertheless, the traditional particle swarm algorithm still exhibits shortcomings in terms of global search, diversity, and convergence. Consequently, scholars have conducted research on enhanced particle swarm algorithms. Zhang et al. [26] proposed a new PSO algorithm with self-adaptive inertia weights and social learning strategies, which improves the algorithm's convergence speed and global searching ability by dynamically adjusting the inertia weights and social learning factors ability. Yang et al. [27] proposed a particle swarm optimisation algorithm based on dual search strategies for dynamic multi-objective optimisation. Two search strategies are designed to update the velocity of each particle, which helps to accelerate the convergence speed and maintain the diversity of the population in a dynamic environment. For convergence improvement, Peng et al. [28] proposed a multi-objective particle optimisation algorithm that incorporates the concepts of shared learning and dynamic congestion distance. The algorithm aims to improve the convergence and diversity of the Pareto front by facilitating knowledge sharing among particles and dynamically adjusting the crowding distance metric. Cheng et al. [29] combined an improved non-dominated sorting strategy and a dynamic weight adjustment mechanism to achieve a balance between global search and local convergence, and utilised constraint-handling techniques to deal with constrained problems. For iterative formulation improvement, Lin et al. [30] combined an effective solution set evaluation method based on divergence statistics and a dynamic weight adjustment strategy to achieve a balance between global search and local convergence, and reduced the complexity of the non-dominated ranking by introducing a polynomial fitting function.

III. Model Building

In an emergency rescue scenario, the precise positioning of UAVs represents a critical design consideration. The lighting area must factor in the geographic scenario's influence and the mission requirements to establish a comprehensive rescue area lighting coverage model that incorporates various elements.

A. 3D Coverage Model of the Lighting System

The 3D coverage model produced by the lighting system primarily comprises a 3D cone model. This model represents a cone created in space by the light emitted from the lighting device, with the lighting device serving as the vertex and the base forming a circle, as illustrated in Fig. 1. In this context, the cross-sectional angle of the lighting system is defined as 60°, and the light's range is dependent on the UAV's altitude, simplifying the subsequent coverage rate calculation.

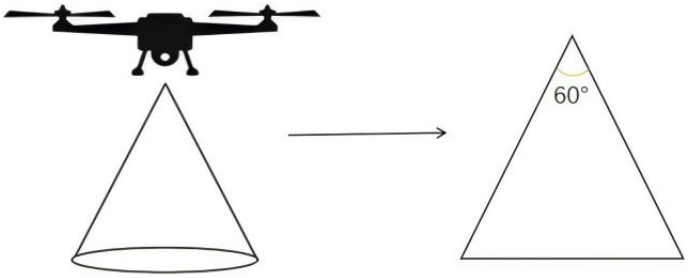


Fig. 1. Three-dimensional conic model of the lighting system and its cross section.

The three-dimensional coverage of the lighting system is initially conceptualized to discretize the volumetric space. However, for practical UAV applications in emergency rescue at specific altitudes, three-dimensional coverage considerations become superfluous. The focus shifts towards two-dimensional plane coverage, which is also subjected to discretization. Within the designated affected area, various characters and obstacles are positioned, necessitating a gridded layout where each grid square denotes a discrete point. Adjustment of the grid step size allows for the manipulation of node positions, facilitating optimal scalability and precision at different accuracy levels. This approach results in the discretization of the two-dimensional plane, as illustrated in Fig. 2, while Fig. 3 visualizes the extent of illumination provided by the two-dimensional lighting system. This model of 3D lighting system coverage effectively simplifies the calculation process, proving highly adaptable for real-world scenarios.

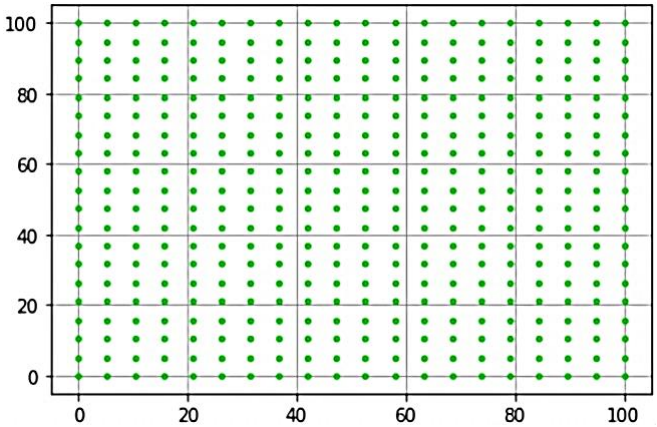


Fig. 2. 2D planar discrete model.

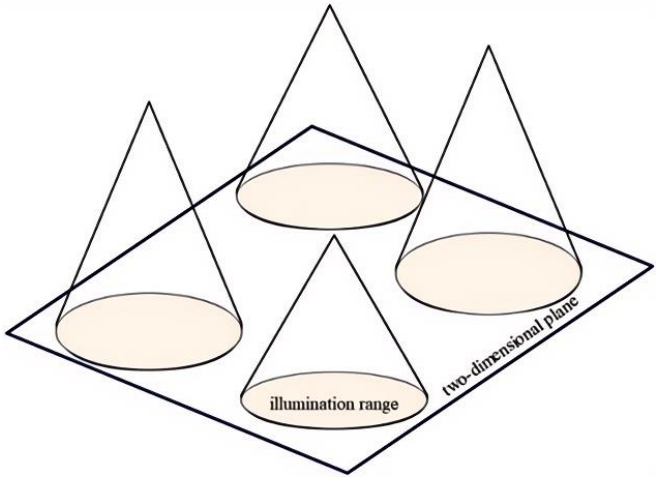


Fig. 3. Multiple unoccupied light ranges in a two-dimensional plane.

B. 3D Geometric Modelling of the Lighting Area

The scheduling of UAVs in rescue scenarios is influenced by lighting requirements and geographic factors. In emergency rescue, the analysis of lighting requirements is centered on providing the necessary illumination to facilitate a range of rescue activities. This analysis necessitates consideration of the specific environmental conditions and site requirements to determine the most suitable lighting solution. It involves detailed examination of the lighting system's type, quantity, placement, intensity, and reliability. Furthermore, emergency rescue lighting demand analysis must address concerns related to energy sources and energy-efficient measures to ensure the stable operation of the lighting system during emergencies.

To meet the specific requirements of emergency rescue operations, the lighting areas are strategically categorized based on the criticality of illumination:

Priority lighting area: This category includes locations crucial for the rescue of individuals or animals, pathways essential for conducting rescue operations, and zones where vital resources are located. It is imperative that the coverage of these areas is meticulously analyzed to ensure at least 90% illumination, serving as a constraint in the optimization challenge.

Other lighting areas: Beyond the priority zones, this category encompasses areas that may contain obstacles, hazardous materials, and miscellaneous regions where maximizing coverage remains a goal.

In outdoor scenarios, the effectiveness of UAV lighting is contingent upon the geographical landscape. For this study, the criteria for adequate coverage in outdoor 3D spatial scenes are defined as follows:

- The affected area must be entirely or partially within the UAV lighting system’s light cone, ensuring clear illumination.
- There must be no obstructions between the UAV system and the target area to avoid shadows and incomplete coverage.

This research concentrates on optimizing the layout of multiple UAVs, deliberately excluding multi-UAV path planning to avoid complexities such as potential collisions. We assume UAVs are deployed to pre-optimized locations, focusing exclusively on layout optimization. This approach allows for a detailed examination and analysis of spatial configuration challenges within multi-UAV systems.

Building on the initial analysis of lighting demands, this study integrates a three-dimensional cone model for the lighting system with a geometric model of the lighting area to establish a comprehensive three-dimensional coverage model, as depicted in Fig. 4. This model not only illustrates the areas covered by light but also applies spatial constraints to calculate precise coverage metrics, enhancing the effectiveness of the UAV lighting strategy in emergency scenarios.

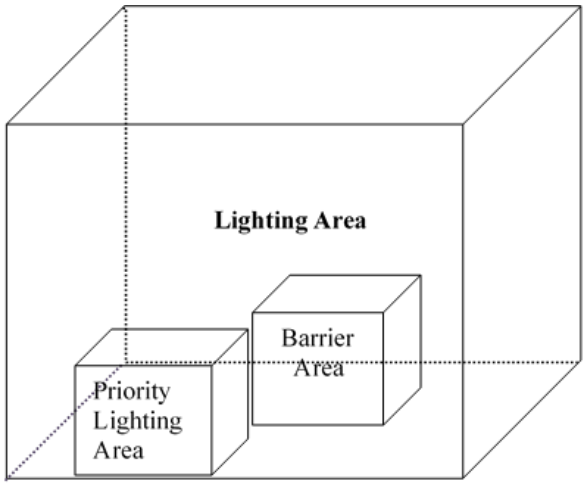


Fig. 4. 3D geometric modelling of lighting areas.

C. Mathematical Modelling

When planning UAV locations, the primary consideration is their ability to adequately illuminate designated areas. In this study, we define a coverage metric as a function with inputs including the 3D coverage model of the lighting system, the 3D geometric model of the illuminated area, the number of UAVs, and the parameters and constraints of the algorithm. The output of this function, the coverage metric, facilitates a comprehensive assessment of the layout's reasonableness, rationality, and effectiveness for the multi-UAV lighting system. The foundational assumptions of the model are:

Uniformity in UAV lighting capabilities: All UAVs are presumed to be equipped with identically capable lighting apparatus, characterized by consistent size, brightness, and irradiation range.

Fixed UAV positions: It is assumed that once each UAV is transported to the designated location and takes off, its position will remain fixed throughout the operation. This implies that our focus is on optimizing the initial layout of UAVs rather than dynamic path planning.

Quantifiable lighting area: The coverage and brightness within the illuminated area are assumed to be quantitatively measurable, allowing for the use of mathematical metrics to evaluate the effectiveness of various lighting configurations, including coverage percentage and brightness uniformity.

Reliability of geographic information: The geographic information employed is presumed accurate and reliable, including details about the area's topography, obstacle locations, and the distribution of critical elements.

No communication interference: It is assumed that there is no communication interference among the UAVs, enabling them to independently execute lighting tasks without the need for real-time communication.

Stable energy supply: A stable energy supply for the UAVs is assumed, ensuring they maintain operational capacity throughout the lighting mission.

1) Definition and meaning of symbols: Identifying constraints and objectives based on the simulated geometric model involves solving a single-objective optimization problem. The variables and parameters used in the model, along with their interpretations, are detailed in Tables I and II.

TABLE I. VARIABLES OF THE THREE-DIMENSIONAL COVERAGE MODEL AND THEIR INTERPRETATION

Variable	Interpretation
$C_i = (x_i, y_i, z_i)$	Geometric position vectors of the UAV
pre_R	Light coverage in priority lighting areas
R_{node}	Node light coverage
R	Total regional light coverage

TABLE II. PARAMETERS OF THE 3D COVERAGE MODEL AND THEIR INTERPRETATION

Parameter	Interpretation
N	Number of UAVs
d	Planar discrete spacing
i	The ith drone
$low_1(x,y,z), up_1(x,y,z)$	Extent of lighting area
$low_2(x,y,z), up_2(x,y,z)$	Range of obstacle locations
$low_3(x,y,z), up_3(x,y,z)$	Scope of Priority Lighting Areas

2) Calculation method of light coverage: Assuming that the lighting region A is a two-dimensional plane, the number of UAVs with the same parameters randomly placed on the region is N, and the coordinates of each UAV are known. The set of UAV nodes is denoted as $C = \{C_1, C_2, \ldots C_N\}$. where $C_i = \{x_i, y_i, z_i\}$ denotes the coordinates of the UAV nodes, and it is also assumed that the light emitted from the illumination system carried by the UAV has a conical top angle of 60°, then the radius of illumination $r = \frac{Z_i}{\sqrt{3}}$.

a) Node set coverage calculation: Assuming that the lighting area A is digitally discretized into $m \times n$ pixels with a discretization spacing of d, the values of m and n are shown in Eq. (1) and (2), respectively.

$$m = (up_1 x - low_1 x) \times d \quad (1)$$

$$n = (up_1 y - low_1 y) \times d \quad (2)$$

Where the coordinate of pixel p is (x, y), then the distance of the pixel from the centre of the drone circle is $d(C_i, p) = \sqrt{(x_i - x)^2 + (y_i - y)^2}$, and defining the event that pixel p is covered by the light from drone i as $r_{i,p}$, the probability of the occurrence of this event, $P_r(r_{i,p})$, is defined as in Eq. (3):

$$P_r(r_{i,p}) = \begin{cases} 1, \iota\phi\ \delta(C_i, p) \le r \\ 0, \text{otherwise} \end{cases} \quad (3)$$

where if $P_r(r_{i,p}) = 1$, it means that pixel point p is covered by the illumination range of UAV i. Therefore, the node coverage is the ratio of the total number of nodes covered by UAVs to the total number of nodes in the region, as in Eq. (4):

$$R_{node} = \frac{\sum_{i=1}^{N} \sum_{p=1}^{m \times n} P_r(r_{i,p})}{m \times n} \quad (4)$$

b) Calculation of area coverage: However, in real-world scenarios, different UAVs may repeatedly cover the same node, requiring the removal of pixel points covered multiple times within the illumination range of multiple UAVs when calculating area coverage. The area coverage of the illuminated region is determined as the ratio of the coverage area of the UAV lighting system to the total area of the illuminated region A, as shown in Eq. (5) and (6):

$$P_{cov(Cov)} = \sum_{p=1}^{m \times n} P_r(r_p) \quad (5)$$

$$\mathrm{R} = \frac{\mathrm{P_{cov(Cov)}}}{\mathrm{m \times n}} \quad (6)$$

where $P_{cov(Cov)}$ denotes the number of unduplicated nodes covered by the UAV lighting system. For any UAV i, if $P_r(r_{i,p}) = 1$, record the node p that is covered by the UAV lighting system at this time and note $P_r(r_p) = 1$. $P_r(r_p) = 1$ indicates that node p is covered by the lighting system.

3) Objective functions and constraints: Coverage problems for UAV node deployment can be divided into two types: area coverage and target coverage. The area coverage focuses on the illumination of the whole area and the target coverage problem focuses on the priority lighting area. The optimisation objective of this problem is to maximise the area coverage as in Eq. (7), P represents the overall coverage rate across the entire region:

$$\mathrm{M}\alpha\xi\ \mathrm{P} \quad (7)$$

Set the constraints of this optimisation problem as:

All UAVs should be located on the surface of the terrain and all inside the area to be illuminated, and secondly, the UAVs must not be inside a three-dimensional obstacle. The UAVs should not be positioned too low, at least three meters above the ground, to avoid interfering with rescue activities, i.e. the coordinates C_i of the UAV satisfy the following Eq. (8):

$$\begin{cases} low_1 x \le x_i \le low_2 x \ or \ up_2 x \le x_i \le up_1 x \\ low_1 y \le y_i \le low_2 y \ or \ up_2 y \le y_i \le up_1 y, i = 1,2,\ldots,N \\ 3 \le z_i \le low_2 z \ or \ up_2 z \le z_i \le up_1 z \end{cases} \quad (8)$$

Avoiding too much repetition of the coverage of different UAV lighting areas, the ratio of the projected distance between any two UAVs on the two-dimensional plane to the sum of their illumination radii is defined to be greater than 0.6, i.e.

satisfying Eq. (9):

$$\sqrt{(x_p - x_q)^2 + (y_p - y_q)^2} > \frac{\sqrt{3}(z_p + z_q)}{5},$$
$$p \neq q, p, q = 1,2,\ldots,N \quad (9)$$

In the disaster area, there are often people or other important things that need to be rescued, so the priority lighting area is established, in which the coverage rate needs to reach more than 90%, defining the coverage rate of the priority lighting area as pre_R, i.e. to satisfy $pre\_R > 0.9$.

IV. Improved Particle Swarm Optimisation Algorithm

This paper employs an enhanced particle swarm algorithm that dynamically adjusts the learning factor and inertia weight factor using diverse methods. It also employs distinct strategies to obtain the individual and global optimal solutions, enhancing search efficiency and accuracy while ensuring a level of stability.

PSO was selected in this study for its simplicity, fast convergence, and strong global search capability. Unlike GA or ACO, PSO does not require complex operations such as crossover or pheromone updating, making it more suitable for time-sensitive emergency rescue scenarios. Additionally, PSO balances exploration and exploitation effectively through adaptive weight adjustments, which aligns well with the requirements of optimizing multi-UAV collaborative lighting layouts.

A. Elementary Particle Swarm Optimisation Algorithm

The core velocity update Eq. (10) and position update Eq. (11) of the elementary particle swarm algorithm can be expressed as follows, respectively:

$$v_{i,d}^{k+1} = wv_{i,d}^{k} + c_1 r_1 \left(x_{i,d,pbest}^{k} - x_{i,d}^{k}\right) + c_2 r_2 \left(x_{d,gbest}^{k} - x_{i,d}^{k}\right) \quad (10)$$

$$x_{i,d}^{k+1} = x_{i,d}^{k} + v_{i,d}^{k+1} \quad (11)$$

where w is the inertia weight, which is used to regulate the global and local search ability of the particle in the search space. c_1 and c_2 are the individual learning factor and the population learning factor, respectively, which are used to regulate the direction and magnitude of the particle updating speed and indicate the magnitude of the learning ability. r_1, r_2 are random numbers of size between 0 and 1, $v_{i,d}^{k}$ is the velocity component of the ith particle in the dth dimension in the kth iteration, $x_{i,d}^{k}$ is the position component of the ith particle in the dth dimension in the kth iteration, $x_{i,d,pbest}^{k}$ is the position component of the individual particle's optimal position of the dth dimension after the kth iteration for the ith particle, and $x_{d,gbest}^{k}$ is the position component of the global optimal position of the dth dimension after the kth iteration. The flow of the optimisation algorithm for particle swarm is shown in Fig. 5.

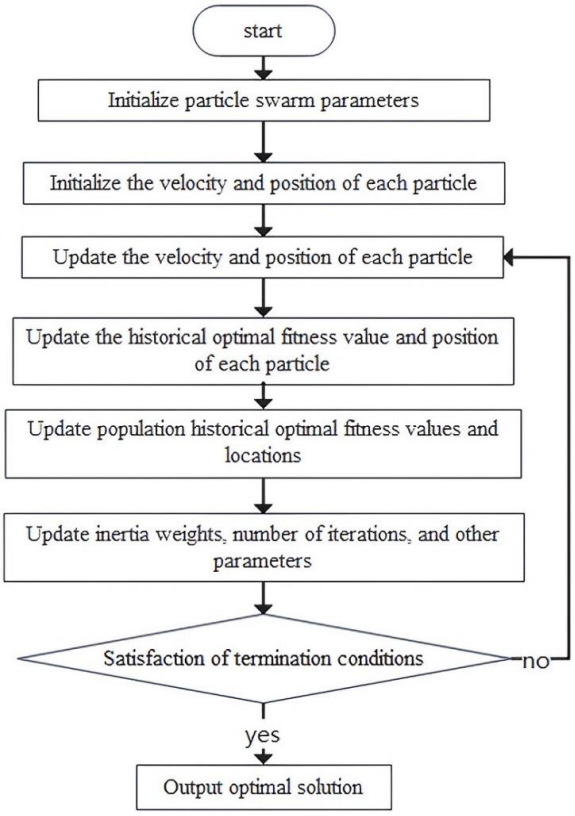


Fig. 5. Particle swarm optimisation algorithm flow.

B. Adjustment of Weights

1) Linearly decreasing inertia weights: In the particle swarm algorithm, the trend of velocity change can be controlled by adjusting the weights w. When w is small, the velocity change of the particle is relatively slow, and it is easy to fall into the local optimal solution; while when w is large, the velocity change of the particle is fast, and it is easy to skip the global optimal solution. Therefore, in order to balance the local search and global search capability, a linearly varying weight strategy is usually adopted, i.e. gradually decreasing from the initial value w_{max} to w_{min}, as shown in Eq. (12), where i is the number of iterations, and i_{max} is the maximum number of iterations.

$$w = w_{max} - (w_{max} - w_{min}) \frac{i}{i_{max}} \quad (12)$$

2) Adaptive inertia weights: The use of adaptive weights in particle swarm optimisation algorithms serves to enhance the algorithm's convergence performance and adaptability. Through dynamic weight adjustments throughout the algorithm's iterations, each particle can more effectively explore the search space and terminate the search more swiftly once the best solution is located.

More specifically, adaptive weights can affect the speed and direction of particle movement towards regions that are more likely to contain the best solution. During the initial stages of the algorithm, larger weights facilitate faster exploration of the search space, whereas in the later stages, smaller weights assist particles in converging more effectively

towards the optimal solution. This dynamic weight adjustment enhances the flexibility and adaptability of the PSO algorithm, enabling it to exhibit improved performance across various problem domains, with the weight changes following the pattern outlined in Eq. (13):

$$w = \begin{cases} w_{min} - (w_{max} - w_{min})\frac{f - f_{min}}{f_{avg} - f_{min}}, f \le f_{avg} \\ w_{max}, f > f_{avg} \end{cases} \quad (13)$$

C. Adjustment of Learning Factors

1) Compression factor method: In particle swarm algorithms, the learning factor is an important parameter used to control the particle speed update. Typically, each particle tracks two historical optimal solutions: the individual historical optimal solution and the neighbourhood historical optimal solution. The learning factors correspond to these two types of historical information, including the individual learning factor c_1 and the social (group) learning factor c_2. Specifically, the individual learning factor controls the extent to which a particle searches near its own historical optimal solution, which makes it more likely that the particle will move in the direction of its own historical optimal solution. The social factor, on the other hand, controls the extent to which a particle searches near its neighbourhood's historical optimal solution, which makes the particle more likely to move towards the neighbourhood's historical optimal solution location. In general, larger individual learning factors facilitated global search, while larger social factors facilitated local search.

In order to effectively control the flight speed of particles and make the algorithm achieve an effective balance between global search and local search, the particle swarm optimisation model that introduces a contraction factor was first proposed by Kennedy and Mendes [19], which employs a compression factor, and this tuning method, by appropriately selecting the parameters, ensures convergence of the particle swarm algorithm and removes the boundary restriction on the speed. The velocity update is formulated as follows in Eq. (14), where the compression factor ϕ is formulated as in Eq. (15).

$$v_{i,d}^{k+1} = \phi[wv_{i,d}^{k} + c_1 r_1\left(x_{i,d,pbest}^{k} - x_{i,d}^{k}\right) + c_2 r_2(x_{d,gbest}^{k} - x_{i,d}^{k})] \quad (14)$$

$$\phi = \frac{2}{\left|2 - (C_1 + C_2) - \sqrt{(C_1 + C_2)^2 - 4(C_1 + C_2)}\right|} \quad (15)$$

2) Asymmetric learning factors: Kennedy proposes an asymmetric learning factor strategy, i.e. the learning factor of each particle is divided into two parts, one part is used for global search and the other part is used for local search. This strategy can make the particles focus more on local search and stop searching faster when the optimal solution is found. The formula of the learning factor is as follows Eq. (16) and Eq. (17), where C_1^{ini} is the initial value of the individual learning factor, C_1^{fin} is the termination value of the individual learning factor, C_2^{ini} is the initial value of the social learning factor, and C_2^{fin} is the termination value of the social learning factor.

$$C_1^k = C_1^{ini} + \left(C_1^{fin} - C_1^{ini}\right)\frac{i}{i_{max}} \quad (16)$$

$$C_2^k = C_2^{ini} + \left(C_2^{fin} - C_2^{ini}\right)\frac{i}{i_{max}} \quad (17)$$

D. Performance Comparison of Different Improvements

In order to verify the effect of the improved algorithm, in the testing process, without setting up the priority lighting area and obstacle area, taking the particle swarm size of 20 and the maximum iteration number of 500, the performance of the algorithm is compared by changing the number of UAVs, the size of the lighting area, and the parameters, interpretation and their specific values of Experiments 1 and 2 are shown in Table III and Table IV, respectively.

TABLE III. THE PARAMETERS, INTERPRETATION AND SPECIFIC VALUES OF EXPERIMENT 1

Parameter	Interpretation	Value
N	Number of UAVs	3
$low_1(x,y,z), up_1(x,y,z)$	Extent of lighting area	(0,0,0),(10,10,10)
$low_1 z, up_1 z$	Altitude range of UAVs	3,7

TABLE IV. THE PARAMETERS, INTERPRETATION AND SPECIFIC VALUES OF EXPERIMENT 2

Parameter	Interpretation	Value
N	Number of UAVs	8
$low_1(x,y,z), up_1(x,y,z)$	Extent of lighting area	(0,0,0),(30,30,30)
$low_1 z, up_1 z$	Altitude range of UAVs	5,10

TABLE V. ALGORITHM PARAMETERS AND THEIR SPECIFIC VALUES

Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
w_{max}	0.9	c_1	2	C_1^{ini}	2.5	C_2^{ini}	1
w_{min}	0.4	c_2	2	C_1^{fin}	4.5	C_2^{fin}	2.25

For ease of recording, the modalities of linearly decreasing inertia weights, adaptive inertia weights, compression factor method, and asymmetric learning factor are defined as ordinal numbers 1, 2, 3, and 4, respectively, and for each experiment, one modality is selected among the weights and the learning factor, respectively, to make up four modalities, i.e. 1-3, 1-4, 2-3, and 2-4. The specific values of the parameters of the algorithm in the experiments are shown in Table V, and the iterations of the improved particle swarm algorithm were carried out using the following parameters, using the value of the coverage calculated in the lighting area as the fitness

function, recording the before and after comparisons of the coverage maps obtained from different combinations under different experiments, as well as the corresponding initial fitness value, the optimal fitness value, the average value of fitness, and the variance obtained from the calculation of the fitness.

A comparison of the 2D planar coverage produced by combinations 1-3, 1-4, 2-3, and 2-4 in Experiment 1 is shown below in Fig. 6-9, and a comparison of the algorithmic iterative graphs for the four improvements is shown in Fig. 10.

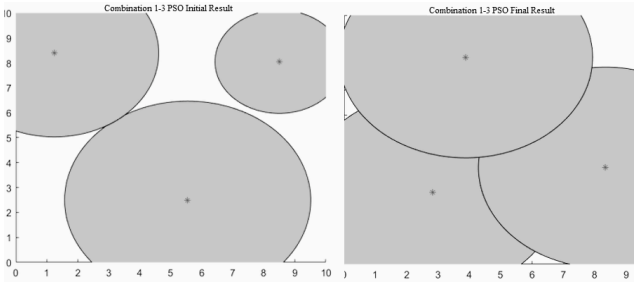


Fig. 6. Experiment 1 Combination 1-3 before and after optimization.

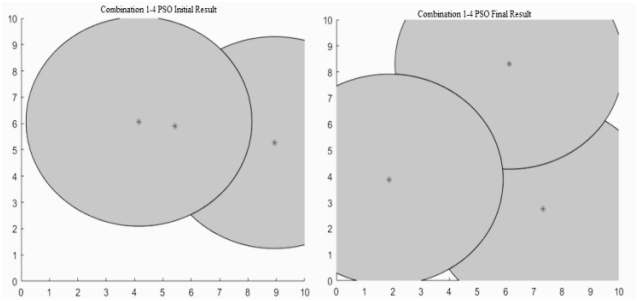


Fig. 7. Experiment 1 Combination 1-4 before and after optimization.

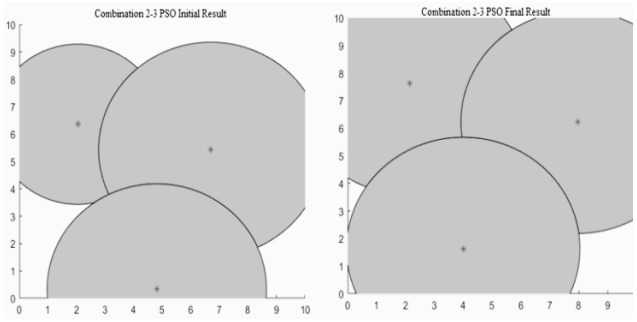


Fig. 8. Experiment 1 Combination 2-3 before and after optimization.

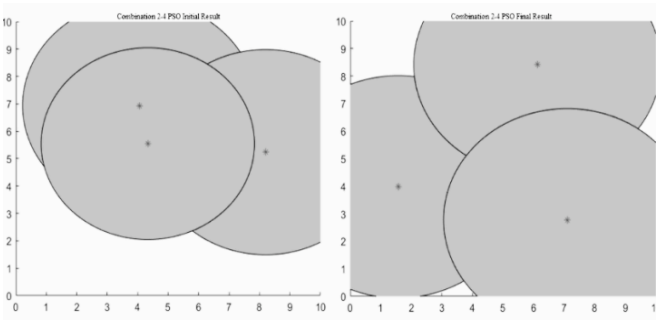


Fig. 9. Experiment 1 Combination 2-4 before and after optimization.

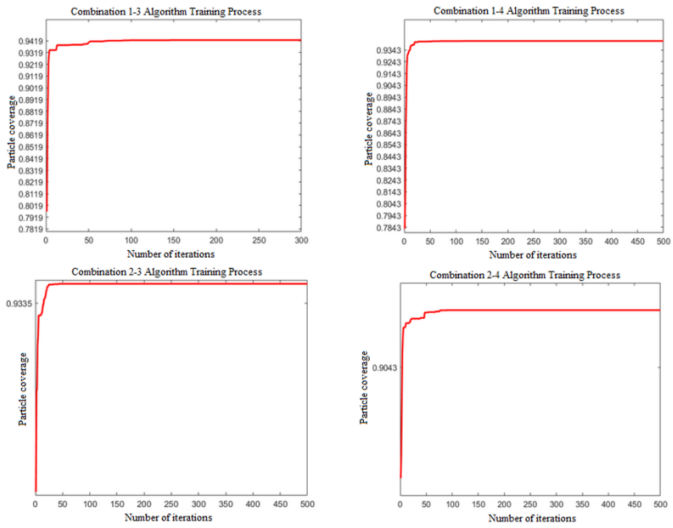


Fig. 10. Experiment 1 Comparison of the iterative process for different combinations.

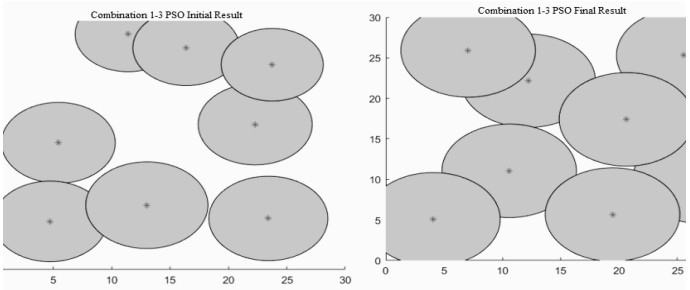


Fig. 11. Experiment 2 Combination 1-3 before and after optimization.

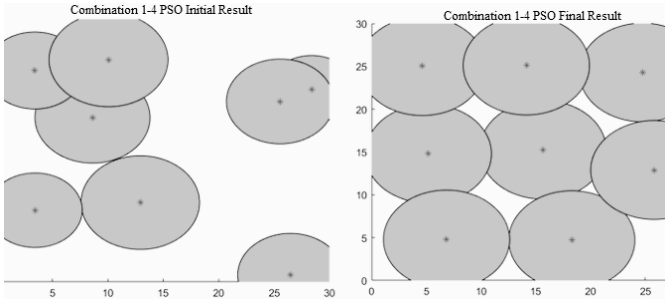


Fig. 12. Experiment 2 Combination 1-4 before and after optimization.

A comparison of the 2D planar coverage produced by combinations 1-3, 1-4, 2-3, and 2-4 in Experiment 2 is shown below in Fig. 11-14, and a comparison of the algorithmic iterative graphs for the four improvements is shown in Fig. 15.

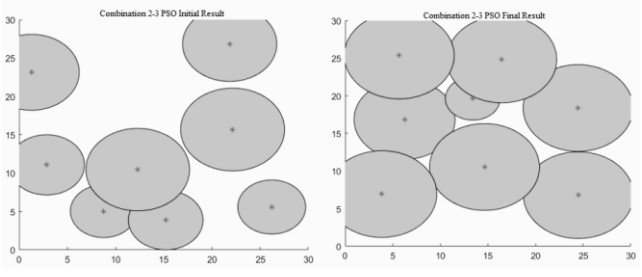


Fig. 13. Experiment 2 Combination 2-3 before and after optimization.

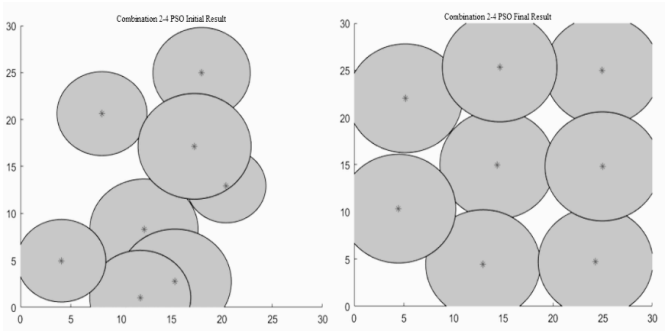


Fig. 14. Experiment 2 Combination 2-4 before and after optimisation.

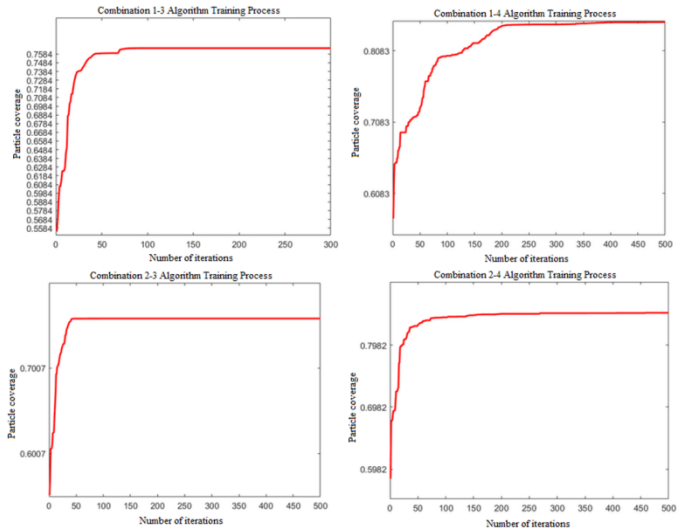


Fig. 15. Experiment 2 Comparison of the iterative process for different combinations.

TABLE VI. PERFORMANCE COMPARISON OF DIFFERENT IMPROVEMENT METHODS

Experiment number	Experimental combinations	Initial fitness	Optimal fitness	Mean fitness	Variance
Experiment 1	1-3	0.751	0.942	0.930	0.0009
	1-4	0.688	0.942	0.912	0.0021
	2-3	0.780	0.939	0.936	0.0008
	2-4	0.687	0.942	0.892	0.0027
Experiment 2	1-3	0.577	0.765	0.747	0.0040
	1-4	0.498	0.849	0.789	0.0067
	2-3	0.497	0.749	0.745	0.0023
	2-4	0.501	0.850	0.801	0.0046

Table VI presents a comparative table of the initial adaptation value, the optimal adaptation value, the average adaptation value, and the variance resulting from adaptation calculations obtained through different experimental improvement methods. From the test results in the table, it is evident that the variance resulting from the experiments is largely consistent and has minimal impact on actual outcomes. However, the approach of utilizing adaptive inertia weights and asymmetric learning factors, specifically the combination of 2-4 in the experiments, yields superior optimisation results compared to other methods. Consequently, a combination of these two enhancement algorithms is employed to address this problem.

V. SIMULATION STUDY

A. Data Preparation

Based on the 3D coverage model of the lighting area, which encompasses the lighting area, priority lighting area, other coverage areas, UAV positions, UAV lighting ranges, and the 2D spatial discrete model, the subsequent 3D model is constructed using the affected area in a real-world scenario, illustrated in Fig. 16.

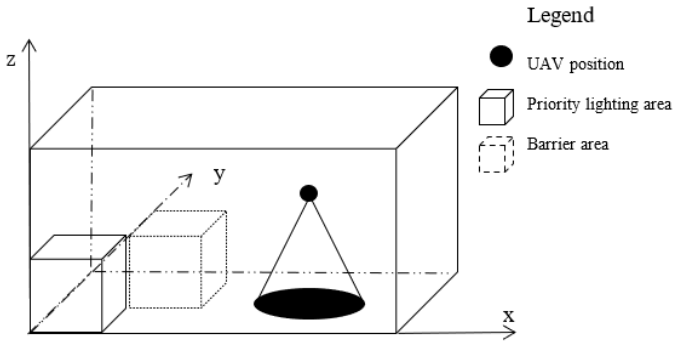


Fig. 16. Lighting area 3D overlay model.

Based on the actual situation of the affected area, the length and width of the lighting area were set at 30 metres and the height at 20 metres, and the specific values of the parameters for the size of the priority lighting area and the obstacle area are shown in Table VII.

TABLE VII. SPECIFIC VALUES FOR THE PARAMETERS OF THE 3D GEOMETRIC MODEL

Parameter	Value	Parameter	Value	Parameter	Value
$low_1(x,y,z)$	(0,0,0)	$low_2(x,y,z)$	(2,2,0)	$low_3(x,y,z)$	(2,0,0)
$up_1(x,y,z)$	(30,30,20)	$up_2(x,y,z)$	(4,4,2)	$up_3(x,y,z)$	(4,2,2)

B. Simulation Results

Assuming that the number of UAVs is 9, the coverage calculation is performed on the lighting areas generated by the nine UAVs, and when the constraints are satisfied, the original UAV positions are entered as the original population to facilitate comparison with the optimised scene. There are nine UAVs in the original scene, and the coverage of the original scene is judged from the calculation of the coverage rate, and the total number of discrete points of 96,100 is obtained by discrete two-dimensional planes with a step size of 0.1, and the lighting of the original scene covers 58,592, with a coverage rate of 0.609, in which the priority lighting area has 900 discrete points, of which 400 are covered, with a coverage rate of 0.444 visible. The initial scene coverage is average and does not meet the desired conditions, of which the illumination coverage area in the 2D plane is shown in Fig. 17.

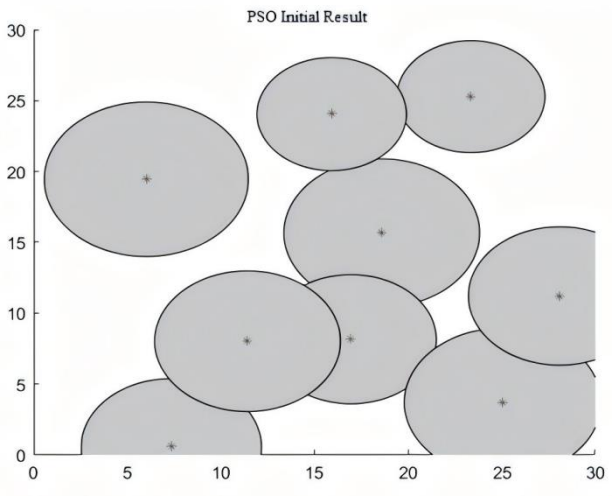


Fig. 17. Particle swarm optimisation algorithm initial coverage results.

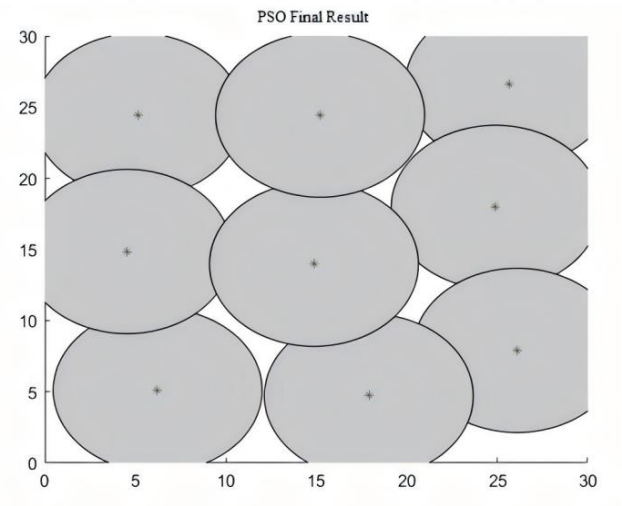


Fig. 18. Optimised 2D coverage map.

Adaptive inertia weights and asymmetric learning factors are chosen to improve the particle swarm algorithm, after the improved particle swarm algorithm is optimised, the coverage nodes are improved compared to the initial scene, by discrete two-dimensional plane with a step size of 0.1, a total of 96,100 discrete points are obtained, and the lighting of the original scene is covered by 86,736, with a coverage rate of 0.903, in which the priority lighting area has a discrete point of 900, of which 900 are covered with a coverage rate of 1. The optimised coverage rate is higher, and at the same time, the UAV positions all satisfy the constraints by not being inside the obstacles, of which the illumination coverage area in the 2D plane is shown in Fig. 18, and the iterative process is shown in Fig. 19. Compared with the simulation experiment without obstacles and priority lighting area, this chapter uses the actual spatial scene for modelling, and the optimisation effect is more realistic. According to the comparison of parameters before and after optimisation, the algorithm is highly effective and adapted to the constraints in emergency rescue, which has a certain degree of universality and can be used in actual rescue with good practical application value.

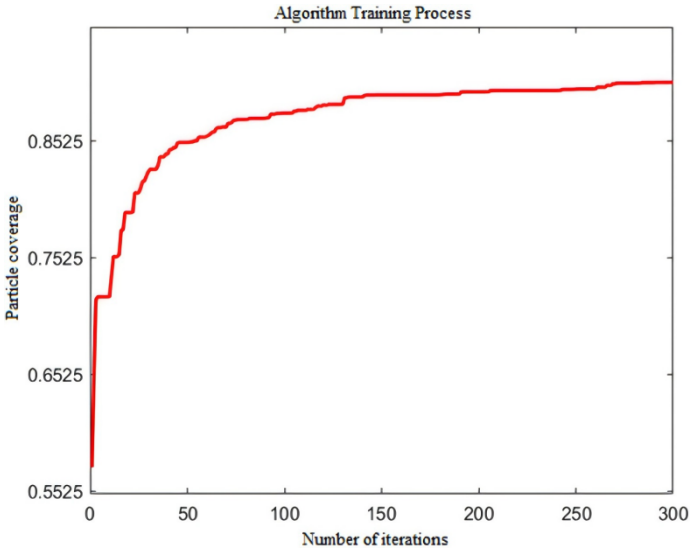


Fig. 19. Algorithmic optimisation process.

C. Effectiveness of the Algorithm

The effectiveness of the improved Particle Swarm Optimization (PSO) algorithm is evident in both the enhanced coverage rate and the optimized UAV deployment strategy.

In the initial UAV layout, the overall coverage rate was 60.9%, with only 44.4% coverage in the priority lighting area. These figures clearly show the inefficiency of the initial deployment, which failed to adequately illuminate critical regions, leaving large portions of the rescue area under-illuminated or overlapping in coverage. After applying the improved PSO algorithm, the overall coverage rate increased significantly to 90.3%, while the priority lighting area achieved a full 100% coverage. This drastic improvement highlights the algorithm's capability to optimize UAV positions effectively, ensuring that crucial rescue areas receive sufficient illumination. This result is crucial for real-world emergency scenarios, where effective lighting directly impacts rescue operations and safety.

The improved PSO algorithm demonstrates superior efficiency in comparison to traditional optimization methods. The adaptive inertia weights allow the algorithm to balance global exploration and local exploitation, preventing premature convergence to suboptimal solutions. Furthermore, the asymmetric learning factors contribute to faster convergence and more accurate UAV positioning, even in complex three-dimensional environments. This is evident from the faster convergence of the algorithm during the simulations, where it reached optimal coverage after fewer iterations compared to traditional methods. For example, in the simulation with 9 UAVs and a step size of 0.1, the optimized algorithm achieved 90.3% coverage with 96,100 discrete points, covering 86,736 points in total. This high coverage, along with improved UAV positioning, ensures that the UAVs avoid obstacles and maximize the effective lighting area.

VI. Conclusion

Collaborative lighting systems involving multiple UAVs can significantly improve the efficiency of accident rescue and

serve as a foundation for informed decision-making during on-site command. The method presented in this study offers an efficient and dependable layout solution for multi-UAV collaborative lighting systems in emergency rescue and related domains. Our research constructs a three-dimensional coverage model of the lighting area based on real-world conditions and addresses the three-dimensional coverage demands of multi-UAV collaborative lighting systems in emergency rescue.

We evaluate the performance of the particle swarm algorithm using various enhancement methods. Adjusting the inertia weights to decrease linearly and modifying the learning factor compression can improve the optimization process, thereby mitigating the particle swarm optimization algorithm's tendency to converge to local optima. This ensures both the speed and accuracy of convergence. Case study results demonstrate that the enhanced particle swarm algorithm effectively optimizes the layout of the multi-UAV lighting system, considering UAV position and coverage area. It guarantees a priority lighting area coverage rate exceeding 90% while ensuring overall lighting coverage.

This study has certain limitations. Firstly, we assumed static UAV positions throughout the operation, without considering dynamic path planning. In real-world emergency scenarios, dynamic factors such as moving obstacles and the need for real-time adjustments to UAV paths may arise, which our current model does not account for. Secondly, the optimization problem addressed in this study is single-objective, focusing solely on maximizing coverage. Important factors such as energy consumption, response time, and the trade-offs between these variables were not incorporated into the model. These assumptions may limit the applicability of the proposed algorithm in more complex rescue scenarios where multiple objectives must be considered simultaneously.

Future research will focus on addressing these limitations. We plan to extend the current model into a multi-objective optimization framework, integrating considerations of energy efficiency, response time, and dynamic path planning. Additionally, the algorithm will be tested in more challenging environments, such as those with dynamic obstacles and changing mission parameters, to validate its robustness and adaptability. Such enhancements will bring the model closer to practical application in real-world emergency rescue missions, where multiple factors influence the success and efficiency of UAV-based lighting systems.

ACKNOWLEDGMENT

This paper was supported by Beijing Natural Science Foundation (Grant No.: 9232022) and Beijing Social Science Foundation Project (Grant No.: 22GLC043).

REFERENCES

[1] Wang, Z. Wu, and Z. Sun, "Optimization of Charging-Station Location and Capacity Determination Based on Optical Storage, Charging Integration, and Multi-Strategy Fusion," J. Green Econ. Low-Carbon Dev., vol. 3, no. 1, pp. 1-14, 2024. https://doi.org/10.56578/jgelcd030101

[2] S. Brischetto and R. Torre, "Preliminary finite element analysis and flight simulations of a modular drone built through fused filament fabrication," J. Compos. Sci., vol. 5, no. 11, p. 293, 2021. https://doi.org/10.3390/jcs5110293

[3] R. Arini, A. D. Putri, W. N. Fadilah, and A. Hasnaoui, "Optimization of Shell and Tube Condenser Effectiveness via PSO Algorithm Coupled with Forced Convection Characterization in Multiphase Systems," Power Eng. Eng Thermophys., vol. 2, no. 4, pp. 188-198, 2023. https://doi.org/10.56578/peet020401

[4] Yahia, A.M., Alkamachi, A. (2024). Design, modeling, and control of tiltable tri-rotors UAV. Journal Européen des Systèmes Automatisés, Vol. 57, No. 3, pp. 841-848. https://doi.org/10.18280/jesa.570323.

[5] H. Hildmann and E. Kovacs, "Using unmanned aerial vehicles (UAVs) as mobile sensing platforms (MSPs) for disaster response, civil security and public safety," Drones, vol. 3, no. 3, p. 59, 2019. https://doi.org/10.3390/drones3030059

[6] A. M. El-Adle, A. Ghoniem, and M. Haouari, "The cost of carrier consistency: Last-mile delivery by vehicle and drone for subscription-based orders," J. Oper. Res. Soc., vol. 75, no. 5, pp. 821-840, 2024. https://doi.org/10.1080/01605682.2023.2210604

[7] H. Li, F. Wang, and Z. Zhan, "Truck and rotary-wing drone routing problem considering flight-level selection," J. Oper. Res. Soc., vol. 75, no. 2, pp. 205-223, 2024. https://doi.org/10.1080/01605682.2023.2185548

[8] Eddine, B.H., Riad, B., Youcef, Z. (2024). Spurious trip rate optimization using particle swarm optimization algorithm. International Journal of Safety and Security Engineering, Vol. 14, No. 1, pp. 63-69. https://doi.org/10.18280/ijsse.140106

[9] J. Liu, J. Wu, and M. Liu, "UAV monitoring and forecasting model in intelligent traffic oriented applications," Comput. Commun., vol. 153, pp. 499-506, 2020. https://doi.org/10.1016/j.comcom.2020.02.009

[10] H. Mhmood, "Design of an Optimized Robust Deadbeat Controller for Roll Motion in Tail-Sitter VTOL UAVs," J. Intell Syst. Control, vol. 3, no. 1, pp. 21-32, 2024. https://doi.org/10.56578/jisc030102

[11] F. Miao, Y. Zhou, and Q. Luo, "A modified symbiotic organisms search algorithm for unmanned combat aerial vehicle route planning problem," J. Oper. Res. Soc., vol. 70, no. 1, pp. 21-52, 2019. https://doi.org/10.1080/01605682.2017.1418151

[12] Y. Cui, X. Fang, G. Liu, and B. Li, "Trajectory optimization of UAV based on Hp-adaptive Radau pseudospectral method," J. Ind. Manag. Optim., vol. 19, no. 1, pp. 675-694, 2022. https://doi.org/10.3934/jimo.2021201

[13] B. Zhou and X. Chen, "A Directional Heuristics Pulse Algorithm for a Two Resources Constrained Shortest Path Problem with Reinitialization," J. Ind. & Manag. Optim., vol. 19, no. 5, pp. 3534-3559, 2023. https://doi.org/10.3934/jimo.2022097

[14] T. Shima and C. Schumacher, "Assigning cooperating UAVs to simultaneous tasks on consecutive targets using genetic algorithms," J. Oper. Res. Soc., vol. 60, no. 7, pp. 973-982, 2009. https://doi.org/10.1057/palgrave.jors.2602617

[15] S. Li, H. Tang, S. He, Y. Shu, T. Mao, J. Li, and Z. Xu, "Unsupervised detection of earthquake-triggered roof-holes from UAV images using joint color and shape features," IEEE Geosci. Remote Sens. Lett., vol. 12, no. 9, pp. 1823-1827, 2015. https://doi.org/10.1109/LGRS.2015.2429894

[16] C. A. Thiels, J. M. Aho, S. P. Zietlow, and D. H. Jenkins, "Use of unmanned aerial vehicles for medical product transport," Air Med. J., vol. 34, no. 2, pp. 104-108, 2015. https://doi.org/10.1016/j.amj.2014.10.011

[17] R. Nagasawa, E. Mas, L. Moya, and S. Koshimura, "Model-based analysis of multi-UAV path planning for surveying postdisaster building damage," Sci. Rep., vol. 11, no. 1, 18588, 2021. https://doi.org/10.1038/s41598-021-97804-4

[18] J. Guo, B. Li, and Y. Ji, "A control parametrization based path planning method for the quad-rotor UAVs," J. Ind. Manag. Optim., vol. 18, no. 2, pp. 1079-1100, 2022. https://doi.org/10.3934/jimo.2021009

[19] X. Ma, D. Wang, N. Zheng, and S. Zhang, "Aggregation and adjustment mechanisms for disaster relief task allocation with uneven distribution," J. Ind. Manag. Optim., vol. 19, no. 3, pp. 1734-1754, 2023. https://doi.org/10.3934/jimo.2022015

[20] C. F. Huang, Y. C. Tseng, and L. C. Lo, "The coverage problem in three-dimensional wireless sensor networks," J. Interconnect. Netw., vol. 8, no. 3, pp. 209-227, 2007.

https://doi.org/10.1142/S0219265907001990

[21] S. N. Alam and Z. J. Haas, "Coverage and connectivity in three-dimensional networks with random node deployment," Ad Hoc Netw., vol. 34, pp. 157-169, 2015. https://doi.org/10.1016/j.adhoc.2014.09.008

[22] P. Gou, B. Guo, M. Guo, and S. Mao, "VKECE-3D: energy-efficient coverage enhancement in three-dimensional heterogeneous wireless sensor networks based on 3D-voronoi and K-means algorithm," Sensors, vol. 23, no. 2, 573, 2023. https://doi.org/10.3390/s23020573

[23] X. Li and X. Y. Zhu, "Design and Testing of Cooperative Motion Controller for UAV-UGV System," Mechatron. Intell Transp. Syst., vol. 1, no. 1, pp. 12-23, 2022. https://doi.org/10.56578/mits010103

[24] Ramesh, N. S. Chandrahas, M. S. Venkatramayya, M. Naresh, P. Talari, D. U. V. D. Prasad, K. S. Kumar, and V. V. Kumar, "Effects of Spacing-to-Burden Ratio and Joint Angle on Rock Fragmentation: An Unmanned Aerial Vehicle and AI Approach in Overburden Benches," Acadlore Trans. Geosci., vol. 2, no. 3, pp. 155-166, 2023. https://doi.org/10.56578/atg020303

[25] J. Kennedy and R. Eberhart, "Particle swarm optimization," Proceedings of ICNN'95 - International Conference on Neural Networks, Perth, WA, Australia, 1995, pp. 1942-1948. https://doi.org/10.1109/ICNN.1995.488968

[26] Y. Zhang, X. Liu, F. Bao, J. Chi, C. Zhang, and P. Liu, "Particle swarm optimization with adaptive learning strategy," Knowl.-Based Syst., vol. 196, 105789, 2020. https://doi.org/10.1016/j.knosys.2020.105789

[27] J. Yang, J. Zou, S. Yang, Y. Hu, J. Zheng, and Y. Liu, "A particle swarm algorithm based on the dual search strategy for dynamic multi-objective optimization," Swarm Evol. Comput., vol. 83, 101385, 2023. https://doi.org/10.1016/j.swevo.2023.101385

[28] G. Peng, Y. W. Fang, W. S. Peng, D. Chai, and Y. Xu, "Multi-objective particle optimization algorithm based on sharing–learning and dynamic crowding distance," Optik, vol. 127, no. 12, pp. 5013-5020, 2016. https://doi.org/10.1016/j.ijleo.2016.02.045

[29] S. Cheng, H. Zhan, and Z. Shu, "An innovative hybrid multi-objective particle swarm optimization with or without constraints handling," Appl. Soft Comput., vol. 47, pp. 370-388, 2016. https://doi.org/10.1016/j.asoc.2016.06.012

[30] Bakirci, M. (2023). A novel swarm unmanned aerial vehicle system: Incorporating autonomous flight, real-time object detection, and coordinated intelligence for enhanced performance. Traitement du Signal, Vol. 40, No. 5, pp. 2063-2078. https://doi.org/10.18280/ts.400524

Analyzing the Impact of Occupancy Patterns on Indoor Air Quality in University Classrooms Using a Real-Time Monitoring System

Sri Ratna Sulitiyanti[1], Muhamad Komarudin[2], F.X Arinto Setyawan[3],
Hery Dian Septama[4], Titin Yulianti[5], M. Farid Ammar[6]
Department of Electrical and Informatics Engineering, University of Lampung, Jl. Soemantri Brojonegoro No. 1, Bandar Lampung, 35145, Indonesia[1, 2, 3, 4, 5, 6]
Environmental Engineering Graduate School, University of Lampung, Jl. Soemantri Brojonegoro No. 1, Bandar Lampung, 35145, Indonesia[2]

***Abstract*—Indoor air quality (IAQ) in universities is of concern because it directly affects students' health and performance. This study presents an IoT-based system for real-time monitoring of IAQ in university classrooms. The system uses MQ-7 and MQ-135 sensors to monitor CO and CO2 pollution parameters. The data is then processed by the ESP32 microcontroller, displayed on the LCD screen, and responded to immediately in the mobile application. The system's real-time monitoring capabilities, data display, and alert mechanism provide valuable insights to improve the classroom environment. The sensors used in the system achieved an accuracy of 97.17% for five people and 93.96% for ten people's scenarios. This study investigates the relationship between human behavior, classroom activities, and occupancy impacts the IAQ. The results show a strong positive correlation between occupancy rates and CO2 levels, indicating the importance of ventilation in densely populated classrooms. The correlation coefficient between the number of students and the CO2 levels is 0.982. This coefficient is remarkably close to 1, indicating a strong positive correlation. In other words, as the number of students in the classroom increases, the CO2 levels also increase significantly. The high correlation coefficient suggests a direct relationship between the number of students and the CO2 levels. This IoT-based system will facilitate a data-driven approach to improving indoor environmental conditions, supporting healthier and more effective learning environments in educational institutions.**

***Keywords*—Indoor air quality; monitoring; pollution; IoT**

I. INTRODUCTION

Indoor air quality is a concern for researchers, especially regarding health. Human activities much more cause air pollution. Industrial gases and vehicle emissions in urban areas are the primary sources of pollutants. Indoor air pollution is a significant health problem because indoor pollutant levels can be 10 to 100 times higher than outdoors [1]. The many sources of hazardous pollutants indoors are risk factors for human health. Common indoor pollutants include particulates, volatile organic compounds, chemical gases, heavy metals, and biological contaminants. The negative impacts of this air pollution are hazardous to human health. This includes the risk of stroke, lung cancer, heart disease, asthma, eye irritation, nasal irritation, and various other diseases [2]. This air pollution has a similar impact to smoking tobacco. From several studies, it is now known that sources of emissions from building materials, furniture, and consumer goods, as well as from the combustion process, are sources of hazardous pollutants in indoor air [3]. Children and women are much more vulnerable because they spend more time indoors. Indoor air quality, especially in cooking areas, can be a source of incomplete combustion. This condition releases toxic gases such as carbon monoxide. Therefore, mitigation is significant to improve indoor air quality. Improving indoor air quality requires various adjustment processes. Avoiding cigarette smoke, using proper ventilation, and choosing low-emission building materials improve indoor air quality [4]. Indoor air pollution in many developing countries has a higher risk than in developed countries. This is mainly due to the combustion of fuels widely used in developing countries. This pollution, therefore, poses significant health risks. Thus, this condition requires more serious mitigation to improve public health in the region [5].

Indoor air quality problems also occur on campuses or colleges. Research on indoor air quality on campus reveals high levels of air pollution. Indoor environmental conditions, such as schools, can have higher levels of pollutants than outdoors. Classrooms and laboratories in various campus environments have poor air quality. The concentration of carbon dioxide (CO2) in classrooms, libraries, and student cafeterias can exceed the standard of 1000 ppm, according from research results conducted by [6], [7]. The increase in carbon dioxide concentration was also followed by high levels of total volatile organic compounds (TVOC) in several locations. Various initiatives are needed to increase campus community awareness of indoor air quality. Monitoring of air quality parameters such as CO and CO2 can be done to raise environmental awareness in educational institutions [8].

The classroom with high pollution will ultimately affect students' health in the long term [9]. Indoor air quality (IAQ) significantly affects university students' satisfaction, study engagement, and cognitive performance. The study found that variations in CO2 levels and other IAQ parameters correlated with students' engagement and memory performance. Poor IAQ conditions can lead to dissatisfaction and reduced cognitive function, ultimately impacting academic performance

[10]. Research by [11] shows the impact of indoor microbial activity on human health. Ventilation systems and clean living habits can reduce indoor pollution levels. The campus building zone or location significantly impacts indoor air quality in campus buildings. Studies [12] and [13] concluded that campus building zones located in places with high pollutant sources can potentially affect the health and performance of occupants.

Other studies, such as those conducted by [14], evaluated indoor air quality in laboratories and classrooms at the Coimbra Health School. The results found that several indoor pollutants exceeded the limits. Students spend a lot of time indoors at school, so their exposure to air pollutants will be affected. However, other studies have also revealed that students and other people are significant sources of indoor pollution in university classrooms. Devices such as computers do not contribute significantly as pollutants [15]. Occupants in schoolrooms affect indoor air quality through respiratory and skin emissions [16]. Considering the above conditions, monitoring and action are needed to improve air quality on campus. Further studies reveal poor indoor air quality in higher education buildings in the UK. Air quality, especially in larger rooms with high occupancy, can be worse, which affects student comfort and performance. Ultimately, the potential health implications of poor indoor air quality can occur in students. Students can experience health conditions such as cough, throat irritation, headaches, eye irritation, and respiratory problems [17], [18]. Ultimately, the study emphasizes the importance of maintaining good IAQ for overall productivity improvement [19], [20].

The above findings underline the need for indoor air quality monitoring. Mitigation measures can be carried out through continuous monitoring and adequate ventilation. Carrying out a monitoring process is expected to ensure healthy indoor air quality conditions for students and staff on university campuses. Monitoring parameters such as temperature, relative humidity, CO, CO2, and PM2.5 levels at various locations on campus needs to be carried out. This monitoring can help management identify campus areas with poor air quality. Knowing the air quality conditions at each location can be used to implement pollution control measures to improve indoor air quality [21]. Monitoring indoor air quality can be done using sensor technology and the internet. Continuous air quality monitoring in university classrooms can be carried out by installing monitoring devices. This policy can create a healthier indoor air quality environment for students, staff, and visitors. The work in [22] implemented a real-time indoor air quality monitoring system in campus classrooms. This study used Internet of Things (IoT)-based technology and Wireless Sensor Networks to detect pollutants such as CO2, dust, temperature, and humidity. Similar work by [23] presents an IoT system for real-time indoor air quality monitoring in a university environment. This system focuses on CO2, CO, and particulate matter pollutants. However, the relationship between room occupancy patterns and air quality is not studied. According to the condition mentioned earlier, conducting further research on indoor air quality is very important.

This research analyzes the relationship between the influence of human behavior, activities, and occupancy patterns on indoor air quality conditions. Furthermore, the results of this research can be used to develop policies to improve university indoor air quality. Excellent or healthy indoor air quality conditions will correlate to enhancing students' health and academic achievement. This research develops an Internet of Things (IoT)-based system for real-time monitoring of indoor environments within university classrooms. The primary objective is to observe and analyze indoor air quality parameters, including carbon monoxide (CO) and carbon dioxide (CO2) levels. These two air quality parameters are chosen since this parameter has a close relation with human occupancy and activities. The designed monitoring system will integrate multiple sensors to monitor air quality metrics. The notification of environmental conditions exceeding the safe limitation will trigger the alarm and be sent by internet to the mobile application. The data collected from monitoring air quality parameters is stored in the database. This research also investigates the relationship between human behavior, classroom activities, and occupancy patterns with indoor air quality. The results show how different factors, such as classroom size, activity intensity, and occupancy, affect indoor air quality parameters. The findings from this research are expected to contribute to policies. Therefore, classroom environments can be optimized for better air quality and occupant comfort. This IoT-based monitoring system provides a data-driven approach to improve indoor environmental conditions. Good indoor air quality can support healthier and more efficient learning environments in educational institutions. This paper is divided into several sections namely, introduction, literature and research framework, methodology, results and discussion and finally a conclusion.

II. Literature and Research Framework

A. Indoor Air Quality (IAQ)

Indoor air quality standards are set to ensure indoor air quality is free from harmful gases or particles above the threshold. Based on research, indoor pollutant levels can be higher than outdoor pollutants. This will certainly affect the health and satisfaction of occupants. Indoor air quality (IAQ) standards and limits are necessary to maintain human health in buildings. This is because long-term exposure to pollutants above the limit can have adverse effects on humans. However, consistent national and regional IAQ standard values have not yet been agreed upon. This lack of regulation has led to uneven IAQ assessments across countries [24], [25]. The World Health Organization (WHO) has established health-based guideline values for several indoor air pollutants. However, these values have not been universally adopted as legally binding standards across countries. This could be due to the complexity and global variability in IAQ regulations. Efforts are underway to develop comprehensive, science-based IAQ guidelines. A study by [26] The research highlights the limitations of indoor environmental quality standards. It also highlights gaps in the evidence supporting thermal, visual, and air quality requirements and urges critical assessment for better standards.

The standards for a healthy room include a temperature range of 18°C-28°C and humidity levels of 40%-60%. This is stated in the Decree of the Minister of Health of the Republic of Indonesia Number 1405/Menkes/SK/XI/2002, addressing

the Office and Industrial Work Environment Health requirements. The average concentration of carbon dioxide indoors is between 250 - 400 parts per million. The concentration of carbon dioxide between 2,000 - 5,000 ppm may affect human health in ways such as headaches, drowsiness, stagnant air, poor focus, an elevated heart rate, and minor nausea [27]. The impact of air quality on human health is also studied by research in [28], [29]. The thresholds for several air quality parameter are presented in Table I.

TABLE I. CRITERIA FOR CHEMICAL POLLUTANT THRESHOLDS

No#	Parameters	Unit	Requirements
1	Sulfur dioxide (SO2)	ppm	0,1 in 24 hour
2	Nitrogen dioxide (NO2)	ppm	0,04 in 24-hour
3	Carbon monoxide (CO)	ppm	9 in 8 hours
4	Carbon dioxide (CO2)	ppm	1000 in 8 hour
5	Lead (Pb)	mg/m3	1,5 in 15 minutes
6	Formaldehyde (HCHO)	ppm	0,1 in 30 minutes
7	Volatile organic compound (VOC)	(VOC)	3 in 8 hours
8	Environmental Tobacco Smoke (ETS)	mg/m3	35 in 24-hour

B. Monitoring Systems

An indoor air quality monitoring system is needed to evaluate and improve air quality within indoor environments. These systems assess air pollutants' conditions, including gases and particulate matter. The monitoring system is used to ensure a healthy and safe indoor environment. Typically, monitoring systems comprise several vital components. Pollutant sensors, data storage, and processing units are implemented into one system. These sensors detect and measure the concentration of air pollutants present. After measurement, the data is processed through a processing unit for storage in the database. Calibration integrates these monitoring systems to effectively and accurately detect pollutants. This brings in the need for calibration of the monitoring devices [30]. Calibration is the most essential step in maintaining the accuracy of sensor readings. The calibration step ensures that data collected for indoor air quality parameters reflects the environmental conditions. Recently, indoor air quality monitoring systems have become so advanced that their capabilities have increased manifold in several prospects [31]. The modern architecture of these monitoring systems integrates wide-area networks, servers, and client devices. This architecture provides extensive and real-time indoor air quality data monitoring and analysis. Such networked architecture allows for the gathering and aggregating data from many sensors and geographic locations. Comprehensive data gives a general view of indoor air conditions. Integrating sensors for such gases as CO and CO2 makes an advanced system able to provide exact and detailed measurements. Real-time air quality monitoring notifies us about changes in air quality in time. Therefore, these results allow immediate corrective actions. Advanced monitoring systems will be able to supply more detailed and accurate data in a timely way to support effective management and control of indoor air quality parameters. Thus, they will contribute toward healthier and more comfortable indoor environments in the long run [32].

III. METHOD

Fig. 1 depicts the proposed design architecture for the IAQ monitoring system in university classrooms. This system consists of three parts: an input block, a processing block, and an output block. The input block utilizes the MQ7 sensor to detect carbon monoxide (CO) and the MQ135 sensor to detect carbon dioxide (CO2). These sensors monitor the air quality parameters in the classroom in real time. The collected data is then transferred to the microcontroller for processing in the processing block. In the processing part, the data is handled by the ESP32, which acts as a microprocessor. The processed data determines the commands to be executed in the output block. A 5V power supply powers the ESP32.

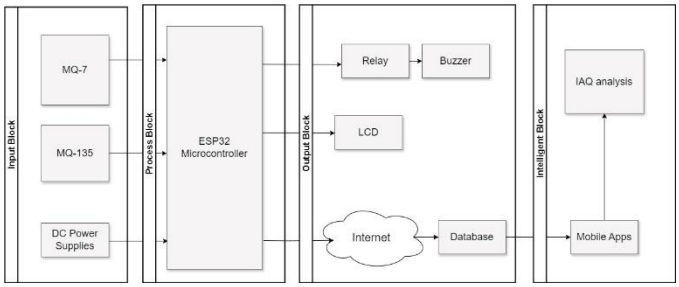


Fig. 1. Proposed system design.

The commands received in the output block trigger appropriate actions based on the processed data. The analysis data is displayed on a 16x2 LCD screen in the enclosure. The system monitors real-time gas levels on the LCD screen and mobile apps. The buzzer is activated when CO and CO2 levels exceed 250 ppm and 1000 ppm, respectively. The buzzer will stop operating when the gas level falls below these limits. This condition indicates that the air quality has returned to a safe level. Data on each sensor's air quality parameters will be stored in a centralized database. This ensures that data can be efficiently accessed and used for various purposes. A summary used by the system's embedded mobile application enables real-time visibility of air quality meters. Users can see their current carbon monoxide, carbon dioxide levels, or any components of interest on their smartphone application. This application offers an easy way to improve air quality comfortably. Besides real-time monitoring, this mobile application can issue warnings when CO2 or CO levels exceed any predetermined limit. These warnings ensure immediate interference to keep the air quality within safety boundaries. The information provided by the application's system contributes to quicker reactions in dangerous situations.

Two different test scenarios were used to gather detailed information. In the first scenario, there were five students in the classroom. In this experiment, CO and CO2 concentrations were recorded at five separate measurements to ensure the accuracy and reliability of the data. This scenario provides the first understanding of how low occupancy rates affect indoor air quality. In the second scenario, there were ten students in the same classroom. Measurements were repeated five times under these conditions to assess the effect of increased

Higher Education for Technology and Innovation (HETI) Project

occupancy on CO and CO2 concentrations. This scenario illustrates the impact of occupancy rates on air quality, particularly regarding CO2 levels.

That is why this monitoring system enhances proficiency in managing indoor air quality. Indoor air quality data was collected and used as a dataset. This enhances, in general, the capability to analyze more and adds value to the monitoring system. The system puts forth a proactive approach toward air quality management. The system is holistic indoor air quality monitoring, highly technological, collecting data, monitoring it in real-time, warning per system deployed, and analysis. This system immediately provides feedback and uses advanced technology to furnish valuable insights and predictions.

IV. Results and Discussions

A. Hardware Implementation

The primary component that sense the air quality parameter include MQ-7 and MQ-135 sensors. Other components are an ESP32 microcontroller, LCD, buzzer, LED indicator, mobile application, and power supply. In this work, an ESP32 microcontroller was chosen due to several GPIO pin arrays for multi-function operations and WiFi connectivity. These modules provide both speed and internet support for system operations and communications. MQ7 and MQ-135 sensors measure CO and CO2, respectively. LCDs instantly read these sensors in real-time to give direct access to the current air quality status. Also, a buzzer is attached to the device as an audible warning system to inform the user if the levels of gases exceed the usually pre-set safety limits. A feature like this would enable the system to monitor the gas level remotely using the mobile application. The setting updates the user with real-time information and contextual air quality status. Thus, users can respond immediately if there is any potential air quality problem inside the classroom. The hardware wiring implementation of the device is shown in Fig. 2, which shows the configuration and integration of these components throughout the system design.

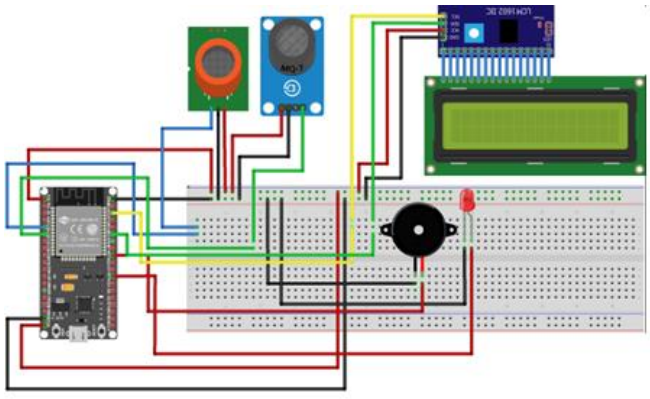


Fig. 2. Hardware wiring implementation.

Fig. 3 depicts the monitoring box implementation. The monitoring system being developed in this study consists of a 9 cm x 15 cm x 10 cm box. The indoor health monitoring system is designed to measure carbon monoxide (CO) with the MQ-7 sensor and carbon dioxide (CO2) concentration with the MQ-135 sensor. This system is integrated with an internet connection and allows real-time data transfer to mobile application users. The Internet and mobile applications enable users to monitor air quality and remotely receive alerts on dangerous levels.

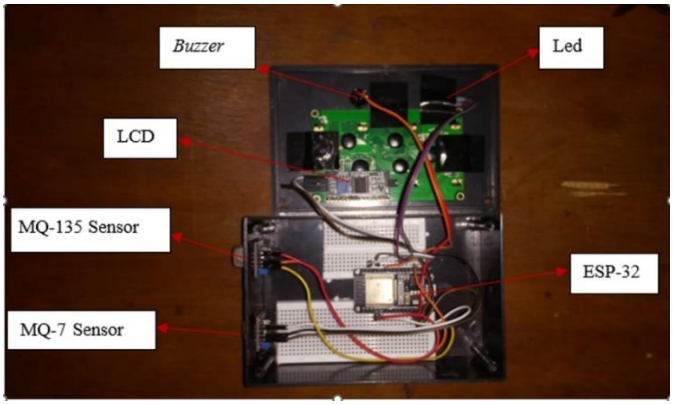


Fig. 3. IAQ hardware implementation.

B. Calibration Process

Once the entire device is assembled, all components are tested and calibrated to produce more accurate results. The method used is linear regression, a statistical technique for comparing sensor outputs with linear characteristics against standard values. In this case, calibration will be performed for the MQ-135 and MQ-7 sensor system readings. The reference instruments are the AZ7752 Instrument and Portable Gas Leak Detector BH-90E. These two devices were chosen as calibrators since they have more accurate results than the low-cost sensors used in this monitoring box. The data resulting from the system calibration are presented in Tables II and III.

TABLE II. MQ-135 Sensors Measurement Process

No.	Calibrator	MQ-135	% error	% accuracy
1	202	203.2	0.59	99.41
2	203	203.2	0.098	99.9
3	204	203.95	0.024	99.98
4	204	204.46	0.22	99.78
5	204	205.3	0.63	99.37
6	204	211.2	3.52	96.48
Average	203.5	205.22	0.847	99.15

TABLE III. MQ-7 Sensors Measurement Process

No.	Calibrator	MQ-7	% error	% accuracy
1	218	216	0.91	99,09
2	218	219	0.45	99.55
3	209	194	7,17	92.83
4	207	193	6.76	93.24
5	208	195	6.25	93.75
6	210	197	6.19	93.81
Average	211.6	202.3	4.62	95.37

The next process is performed using linear regression with Eq. (1). The Y represents the value obtained from the reference instruments Az 7752 and BH-90E as calibrators, and the X denotes the design device sensors' output value.

$$Y = a + bX \quad (1)$$

Fig. 4 and Fig. 5 show the correlation calculation between the readings from the designed device and the reference instrument or calibrators using the linear regression equation.

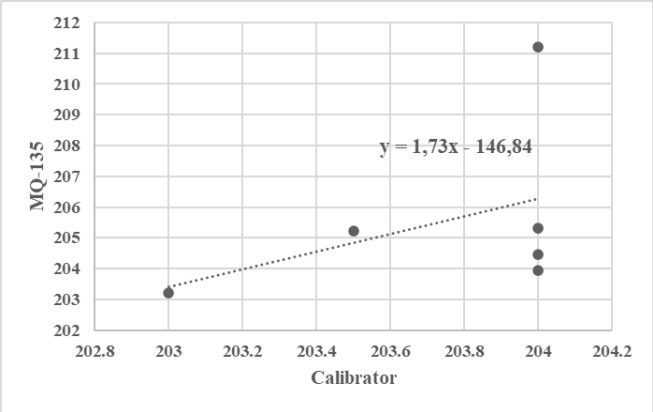


Fig. 4. MQ-135 linear regression graph.

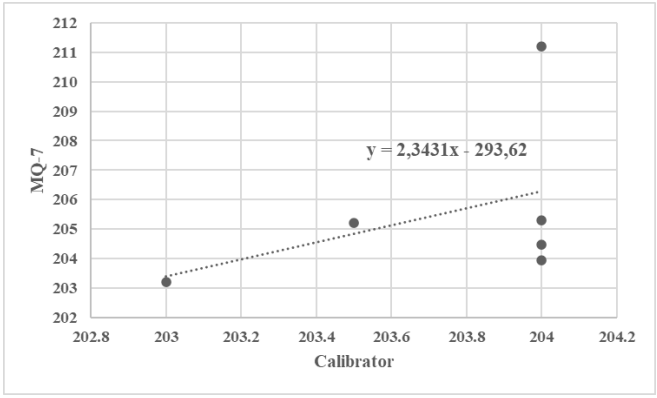


Fig. 5. MQ-7 Linear regression graph.

The linear regression results for the MQ-135 sensor are represented by the equation y = 1.73x – 146.84, and for the MQ-7 sensor, y = 2.3431x – 293.62. These regression equations provide calibration adjustments to enhance the accuracy of the raw sensor readings. By implementing these equations into the source code, the sensors' measurements can be corrected in real-time. This setting allows the system to compensate for any inaccuracies inherent in low-cost sensors. Consequently, the monitoring box that relies on low-cost sensors can deliver more precise and reliable air quality data.

TABLE IV. IAQ TESTING RESULTS

No	number of students	CO		CO2		% error CO2	% accuracy CO2	Notification	
		Sensor	Calibrator	Sensor	Calibrator			Buzzer	Apps
1	5	0	0	211.2	204	3.53	96.47	*OFF*	*OFF*
2	5	0	0	211.2	204	3.53	96.47	*OFF*	*OFF*
3	5	0	0	210.1	204	2.99	97.01	*OFF*	*OFF*
4	5	0	0	210.1	204	2.99	97.01	*OFF*	*OFF*
5	5	0	0	215.15	210	2.45	97.55	*OFF*	*OFF*
6	5	0	0	215.02	210	2.39	97.61	*OFF*	*OFF*
7	5	0	0	211.1	207	1.98	98.02	*OFF*	*OFF*
8	5	0	0	211.17	207	2.01	97.99	*OFF*	*OFF*
9	5	0	0	211.17	207	2.01	97.99	*OFF*	*OFF*
10	5	0	0	216.1	207	4.40	95.60	*OFF*	*OFF*
11	5	0	0	250.3	255	1.84	98.16	*OFF*	*OFF*
12	10	0	0	250.3	255	1.84	98.16	*OFF*	*OFF*
13	10	0	0	254.32	262	2.93	97.07	*OFF*	*OFF*
14	10	0	0	260.1	273	4.73	95.27	*OFF*	*OFF*
15	10	0	0	260.15	277	6.08	93.92	*OFF*	*OFF*
16	10	0	0	264.1	281	6.01	93.99	*OFF*	*OFF*
17	10	0	0	266.15	298	10.69	89.31	*OFF*	*OFF*
18	10	0	0	264.17	284	6.98	93.02	*OFF*	*OFF*
19	10	0	0	263.05	279	5.72	94.28	*OFF*	*OFF*
20	10	0	0	267.2	309	13.53	86.47	OFF	OFF

C. Discussion

The testing results include a range of air quality measurements, including information on the number of students attending classes, carbon monoxide (CO) levels, carbon dioxide (CO2), and monitoring equipment. The CO2 parameters can increase dramatically due to increased human activity and respiration. Results from both scenarios are considerably detailed in Table IV. By comparing the data from these two scenarios, insights can be gained into how occupancy rates affect indoor air quality, particularly CO2 levels. Since CO concentration from the sensor and calibrator shows the exact measurement, the calculation of error and accuracy is not done for the CO. According to Table IV, the average accuracy of CO2 level is 95.57%, with an error of 4.43%.

The average values acquired from this measurement under two scenarios are summarized in Table V. The first scenario was performed on a workday with five college students in a 7 x 7-meter room. The average carbon dioxide (CO2) measurement was 212.23 ppm, with carbon monoxide (CO) degrees measured at zero ppm. The average CO2 concentration in this condition was relatively low, which can be attributed to the confined wide variety of occupants in the given area. In the second stage of experiments, with the room occupied with the aid of ten people, the CO2 awareness increased to an average of 259.98 ppm. This rise in CO2 stages is prompted by numerous factors, including elevated occupants, the room's dimensions, and insufficient air stream inside the area.

TABLE V. AVERAGE IAQ TESTING RESULTS

No.	Condition	CO_2 (ppm)	% error	% accuracy
1.	Five people	212.23	2.83	97.17
2.	Ten people	259.98	5.12	93.96

The findings suggest that a 7 x 7-meter room, when occupied with the aid of 10-15 people for prolonged intervals, remains within a secure variety of air quality. This conclusion is supported by the records supplied in Table V, which show that CO2 ranges and other readings range with adjustments in occupancy. The environment remains possible in these situations.

We use the Pearson correlation formula in Eq. (2) to calculate the correlation coefficient between the number of students and the CO2 levels. According to the correlation analysis, the correlation coefficient is 0.982.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (2)$$

This coefficient is remarkably close to 1, indicating a strong positive correlation. In other words, as the number of students in the classroom increases, the CO2 levels also increase significantly. The high correlation coefficient suggests a direct relationship between the number of students and the CO2 levels. This is expected because humans exhale CO2, and in a closed environment like a classroom, higher occupancy will naturally lead to higher CO2 concentrations. Elevated CO2 levels can impact cognitive function and overall comfort. Indoor CO2 concentrations should be below 1000 ppm to ensure a healthy and productive environment. The data indicates that classrooms with more students will likely have higher CO2 levels. This is potentially exceeding recommended limits if not correctly managed. The strong correlation highlights the importance of adequate ventilation systems in classrooms. Schools should ensure sufficient ventilation rates to maintain CO2 levels within safe limits, particularly in densely occupied classrooms.

The sample size of air quality and the number of classrooms may constrain this study's findings. This is due to limited resources and time for air quality data collection, which may not be feasible. In the future, this indoor air quality monitoring box can be enhanced by adding more sensors to measure different air quality parameters, such as temperature, humidity, and particulate matter. Since the electronic components and sensors are low-cost, the device can be produced in larger quantities for use in other classrooms. All monitoring devices may store the data in the cloud-based database. Although several studies have implemented the idea of monitoring devices. The work in study [33] and study [34] already studied and enabled comprehensive indoor air quality monitoring in classrooms. The work uses a network of intelligent IoT sensors. It may continuously track parameters such as indoor/outdoor temperatures, relative humidity, and CO2 levels, facilitating centralized control of natural ventilation. However, the idea can be extended by using machine learning to analyze the classroom occupancy level and the air quality index. Therefore, the system allows for real-time data processing that is essential for assessing infection risks and optimizing air quality.

V. CONCLUSION

This paper aims to design and test an IoT-based IAQ monitoring system in a university classroom. The monitoring device allows for real-time monitoring of carbon monoxide (CO) and carbon dioxide (CO2) levels. Real-time monitoring, data display, and warnings signaled by the system are important to improve the classroom environment. Accordingly, the system achieved accuracy in sensing at a rate of 97.17% for five people and 93.96% for ten people's scenarios. The Pearson correlation formula compares the number of people's occupancy and the classroom. The Pearson correlation coefficient between the number of students with CO2 levels was 0.982. This result is close to 1, which recommends a strong positive correlation. That is, with increasing numbers of students within the classroom, drastically increasing levels of CO2 are realized. The high correlation coefficient suggests the increased number of students and CO2 levels correspondingly. Continuous monitoring of IAQ and effective ventilation techniques are essential for maintaining a healthy indoor environment in university classrooms. Further research is required to establish how other environmental parameters can be studied across various indoor environments. Another advanced research can integrate sensed pollution data into machine learning for predictive analytics for these IAQ parameters.

ACKNOWLEDGMENT

We extend our deepest gratitude to the Higher Education for Technology and Innovation (HETI) Project for their generous support and funding of our research through the

Domestic Scheme of Research, Innovation, and Collaboration (Riset Inovasi dan Kolaborasi Skema Domestik) 2023.

REFERENCES

[1] J. Jaquarta, "Indoor Air Quality in Homes," IJETI, Vol. 4, No. 3, Sep. 2023, DOI: 10.51626/IJETI.2023.04.00059.

[2] E. Ridolo, A. Pederzani, A. Barone, M. Ottoni, M. Crivellaro, and F. Nicoletta, "Indoor Air Pollution and Atopic Diseases: A Comprehensive Framework," Explor Asthma Allergy, Vol. 2, No. 3, pp. 170–185, May 2024, DOI: 10.37349/EAA.2024.00038.

[3] T. S. Oboronova, M. G. Prodanchuk, S. P. Chumak, O. O. Bobyliova, N. V. Kurdil, and O. L. Pereguda, "Indoor Air Quality as an Important Determinant of Health and Apriority Task in the Field of Human Environmental Hygiene," Ohnpu, Vol. 58, No. 1, pp. 7–18, Jun. 2023, DOI: 10.33273/2663-9726-2023-58-1-7-18.

[4] A. Addisu, "Indoor Air Pollution," In Environmental Sciences, Vol. 9, M. Eyvaz, A. Albahnasawi, and M. Y. D. Alazaiza, EDS., Intechopen, 2023. DOI: 10.5772/Intechopen.110587.

[5] S. Mamuya and J. Bachwenkizi, "A Review on Indoor Air Pollution in Developing Countries," In Reference Module in Biomedical Sciences, Elsevier, 2024, P. B9780323999670002738. DOI: 10.1016/B978-0-323-99967-0.00273-8.

[6] N. A. Khalil and G. M. I. Kamoona, "The Effect of Indoor Air Quality in University Classrooms on the Immunity of its Occupants," IJSDP, Vol. 17, No. 8, pp. 2453–2461, Dec. 2022, DOI: 10.18280/IJSDP.170813.

[7] J. F. Tsai and C.-C. Lin, "Air Quality Investigation and Research for Various Types of Student-Oriented Indoor Environments on a University Campus," E3s Web of Conf., Vol. 396, P. 01031, 2023, DOI: 10.1051/E3SCONF/202339601031.

[8] A. Mundackal and V. M. Ngole-Jeme, "Evaluation of Indoor and Outdoor Air Quality in University Academic Buildings and Associated Health Risk," International Journal of Environmental Health Research, Vol. 32, No. 5, pp. 1076–1094, May 2022, DOI: 10.1080/09603123.2020.1828304.

[9] S.-S. Kim, G.-Y. Hong, D.-K. Kim, S.-S. Kim, and W.-H. Yang, "Estimating Personal Exposures to Air Pollutants in University Students Using Exposure Scenario," Journal of Environmental Science International, Vol. 22, No. 1, pp. 47–57, Jan. 2013, DOI: 10.5322/JES.2013.22.1.47.

[10] Q. Carton, S. De Coninck, J. Kolarik, and H. Breesch, "Assessing the Effect of a Classroom IEQ on Student Satisfaction, Engagement and Performance," E3s Web of Conf., Vol. 396, P. 01052, 2023, DOI: 10.1051/E3SCONF/202339601052.

[11] H. Zhao, R. Du, Y. Liu, D. Wang, and Y. Li, "Assessing Indoor PM2.5 Microbial Activity in a University Campus Environments in Beijing," Building and Environment, Vol. 246, P. 111003, Dec. 2023, DOI: 10.1016/J.BUILDENV.2023.111003.

[12] G. Erlandson, S. Magzamen, E. Carter, J. L. Sharp, S. J. Reynolds, and J. W. Schaeffer, "Characterization of Indoor Air Quality on a College Campus: A Pilot Study," IJERPH, Vol. 16, No. 15, P. 2721, Jul. 2019, DOI: 10.3390/IJERPH16152721.

[13] S. Bellizzi, C. M. Panu Napodano, G. Pichierri, and N. Muthu, "Indoor Air Quality: Persisting Global Issue that Impacts Students' Performance and Health," Indoor and Built Environment, Vol. 30, No. 9, pp. 1587–1588, Nov. 2021, DOI: 10.1177/1420326X211034897.

[14] F. Moreira, A. Ferreira, J. P. Figueiredo, and R. Ferreira, "Indoor Air Quality, Ultrafine Particles in Laboratories and Classrooms of Coimbra Health School," in Proceedings of the 1st International Conference on Water Energy Food and Sustainability (ICOWEFS 2021), J. R. Da Costa Sanches Galvão, P. S. Duque De Brito, F. Dos Santos Neves, F. G. Da Silva Craveiro, H. De Amorim Almeida, J. O. Correia Vasco, L. M. Pires Neves, R. De Jesus Gomes, S. De Jesus Martins Mourato, and V. S. Santos Ribeiro, EDS., Cham: Springer International Publishing, 2021, pp. 537–547. DOI: 10.1007/978-3-030-75315-3_58.

[15] G. Dabanlis, G. Loupa, D. Liakos, and S. Rapsomanikis, "The Effect of Students, Computers, and Air Purifiers on Classroom Air Quality," Applied Sciences, Vol. 12, No. 23, P. 11911, Nov. 2022, DOI: 10.3390/APP122311911.

[16] M. Kruza and N. Carslaw, "How do Breath and Skin Emissions Impact Indoor Air Chemistry?," Indoor Air, Vol. 29, No. 3, pp. 369–379, May 2019, DOI: 10.1111/INA.12539.

[17] B. Oroji, A. Sadighzadeh, E. Solgi, and M. S. Oliaei, "Impact of Air Quality on Students' Behavior in the Educational Centers," Air Qual ATMOS Health, Vol. 14, No. 6, pp. 793–806, Jun. 2021, DOI: 10.1007/S11869-021-00979-Y.

[18] F. Aliakbari, S. T. Moghadam, and P. Lombardi, "Indoor Air and Light Quality Assessment in a University Campus Classroom," SACE, Vol. 30, No. 1, pp. 163–182, Jun. 2022, DOI: 10.5755/J01.SACE.30.1.30328.

[19] S. Lama, C. (Charlie) Fu, and A. Lee, "Indoor Air Quality (IAQ) Evaluation of Higher Education Learning Environments," J. of SMAR. Build. and Constr. Technol., Vol. 4, No. 1, pp. 1–14, Jan. 2022, DOI: 10.30564/JSBCT.V4I1.4042.

[20] K. Daxini, T. Turakhia, R. Iyer, and A. Chhabra, "Assessment of Ambient Air Quality of a College Campus," in 2020 IEEE India Geoscience and Remote Sensing Symposium (INGARSS), Ahmedabad, India: IEEE, Dec. 2020, pp. 193–196. DOI: 10.1109/INGARSS48198.2020.9358939.

[21] W. Liu, Y. He, and Z. Liu, "Indoor Pollution Control based on Surrogate Model for Residential Buildings," Environmental Pollution, Vol. 346, P. 123638, Apr. 2024, DOI: 10.1016/J.ENVPOL.2024.123638.

[22] A. A. Hapsari, A. I. Hajamydeen, D. J. Vresdian, M. Manfaluthy, L. Prameswono, and E. Yusuf, "Real Time Indoor Air Quality Monitoring System based on IoT Using MQTT and Wireless Sensor Network," in 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS), Kuala Lumpur, Malaysia: IEEE, Dec. 2019, pp. 1–7. DOI: 10.1109/ICETAS48360.2019.9117518.

[23] P. Asthana and S. Mishra, "IoT Enabled Real Time Bolt Based Indoor Air Quality Monitoring System," in 2018 International Conference on Computational and Characterization Techniques in Engineering & Sciences (CCTES), Lucknow, India: IEEE, Sep. 2018, pp. 36–39. DOI: 10.1109/CCTES.2018.8674076.

[24] O. Toyinbo Et Al., "Open Database for International and National Indoor Environmental Quality Guidelines," Indoor Air, Vol. 32, No. 4, Apr. 2022, DOI: 10.1111/INA.13028.

[25] J. Saffell and S. Nehr, "Improving Indoor Air Quality Through Standardization," Standards, Vol. 3, No. 3, pp. 240–267, Jul. 2023, DOI: 10.3390/Standards3030019.

[26] A. Mahdavi and C. Berger, "Toward a Critical Assessment of Indoor Environmental Quality Standards," App, Vol. 38, pp. 5–10, Dec. 2022, DOI: 10.14311/App.2022.38.0005.

[27] J. William and A. H. Rosenfeld, "Estimates of Improved Productivity and Health From Better Indoor Environments," International Journal of Air Quality and Climate, Vol. 7, 1997.

[28] P. Kapalo, F. Domniţa, C. Bacoţiu, and N. Spodyniuk, "The Impact of Carbon Dioxide Concentration on the Human Health - Case Study," Journal of Applied Engineering Sciences, Vol. 8, No. 1, pp. 61–66, May 2018, DOI: 10.2478/Jaes-2018-0008.

[29] P. Wolkoff, "Indoor Air Humidity, Air Quality, and Health – An Overview," International Journal of Hygiene and Environmental Health, Vol. 221, No. 3, pp. 376–390, Apr. 2018, DOI: 10.1016/J.IJHEH.2018.01.015.

[30] T. Araújo, L. Silva, A. Aguiar, and A. Moreira, "Calibration Assessment of Low-Cost Carbon Dioxide Sensors Using the Extremely Randomized Trees Algorithm," Sensors, Vol. 23, No. 13, P. 6153, Jul. 2023, DOI: 10.3390/S23136153.

[31] I. Demanega, I. Mujan, B. C. Singer, A. S. Anđelković, F. Babich, and D. Licina, "Performance Assessment of Low-Cost Environmental Monitors and Single Sensors Under Variable Indoor Air Quality and Thermal Conditions," Building And Environment, Vol. 187, P. 107415, Jan. 2021, DOI: 10.1016/J.Buildenv.2020.107415.

[32] M. Komarudin, H. D. Septama, T. Yulianti, A. Yudamson, M. Pratama, and T. P. Zuhelmi, "Air Quality Monitoring Device for Smart Health Solution During Covid-19 Pandemic," in 2021 International Conference on Converging Technology in Electrical and Information Engineering (ICCTEIE), Bandar Lampung, Indonesia: IEEE, Oct. 2021, pp. 78–81. DOI: 10.1109/ICCTEIE54047.2021.9650638.

[33] N. Barros Et Al., "Schoolair: A Citizen Science IoT Framework Using Low-Cost Sensing for Indoor Air Quality Management," Sensors, Vol. 24, No. 1, P. 148, Dec. 2023, DOI: 10.3390/S24010148.

[34] A. Zivelonghi and A. Giuseppi, "Smart Healthy Schools: An IoT-Enabled Concept for Multi-Room Dynamic Air Quality Control," Internet of Things and Cyber-Physical Systems, Vol. 4, pp. 24–31, 2024, DOI: 10.1016/J.Iotcps.2023.05.005.

Active Semi-Supervised Clustering Algorithm for Multi-Density Datasets

Walid Atwa[1], Abdulwahab Ali Almazroi[2], Eman A. Aldhahr[3], Nourah Fahad Janbi[4]
Department of Information Technology-College of Computing and Information Technology at Khulais, University of Jeddah, Jeddah, Saudi Arabia[1, 2, 4]
Department of Computer Science and Artificial Intelligence-College of Computer Sciences and Engineering, University of Jeddah, Jeddah, Saudi Arabia[3]

***Abstract*—Semi-supervised clustering with pairwise constraints has been a hot topic among researchers and experts. However, the problem becomes quite difficult to manage using random constraints for clustering data when the clusters have different shapes, densities, and sizes. This research proposes an active semi-supervised density-based clustering algorithm, termed "ASS-DBSCAN," designed specifically for clustering multi-density data. By integrating active learning and semi-supervised techniques, ASS-DBSCAN enhances traditional clustering methods, allowing it to handle complex data distributions with varying densities more effectively. This research provides two major contributions. The first contribution of this research is to analyze how to link constraints (including that must be linked and ones that should not be linked) that will be utilized by the clustering algorithm. The second contribution made by this research is the ability to add multiple density levels to the dataset. We perform experiments over real datasets. The ASS-DBSCAN algorithm was evaluated against existing state-of-the-art system for various evaluation metrics in which it performed remarkably well.**

***Keywords*—Semi-supervised clustering; pairwise constraints; multi-density data; active learning**

I. Introduction

In data mining, clustering algorithms are used to divide data into multiple groups based on selected similarity metrics [1]. The rate at which Web data is increasing also called Big Data, clustering algorithms are expected to death with 1) scalability, 2) noise, 3) multidimensional data, 4) discovering clusters with arbitrary shapes, and 5) least dependency of domain knowledge to establish input parameters [2].

Clustering algorithms are categorized into several categories based on the requirements of the problem given. This includes 1) Partitional, 2) grid-based, 3) hierarchical and 4) density-based clustering algorithms [3]. From the list, density-based are suitable for finding clusters based on their unique size and shape. It does this by focusing on the density of the clusters in the region as opposed to clusters falling where the density is low [4-7].

The DBSCAN algorithm is one of the most popular used techniques in relation to this category of clustering algorithms that carries forward all advantages of density-based clustering family [8]. It works by calculating the total number of points (called *Eps*) around the point in a region. Points having density above the specified threshold value (called *MinPts*) are referred to as Core points, while others that do not meet these criteria are referred to as Noise points [9].

While this clustering technique is able to identify cluster regions having different shapes or sizes, it cannot handle data that contains clusters having points with different densities. This technique is only able to identify them as a cluster if the points are separated by sparse regions.

For example, Fig. 1(a) shows the dataset contains two cluster groups, sparse cluster C_1 (red points) and dense cluster C_2 (green points), and noise data (black points). When using large *Eps*-distance, DBSCAN algorithm can find a sparse cluster successfully. However, as shown in Fig. 1(b), the algorithm merges the dense cluster with the adjacent noise points. While using small *Eps*-distance, as shown in Fig. 1(c), the DBSCAN algorithm can assist in finding dense cluster C_2. On the other hand, it incorrectly marks the sparse clusters are noise points instead due to the core object condition. As a result of its reliance on set global parameters, DBSCAN is unable to discover clusters with variable densities.

Moreover, most existing density-based algorithms are unsupervised learning methods that can't utilise the available label information such as background knowledge [10] [11]. However, semi-supervised clustering algorithms can be considered as they utilize existing knowledge to enhance the clustering process. However, if these constraints are improperly selected the clustering performance will be affected [12]. Active learning was originally proposed to select the most important labeled data to generate pairwise constraints. However, most of the data in real-world applications is unlabeled [13-15].

To address the issue highlighted above, this is an enhanced version of DBSCAN algorithm that follows an active semi-supervised clustering approach. This new algorithm is named as ASS-DBSCAN which is capable of clustering multi-density data with active pairwise constraints. By quantitively analyzing the data using statistical techniques, the research was able to split the dataset into various density groups. Then, generate a set of active pairwise constrints from existing groups and compute their *Eps*. Finally, expand the clusters using selected active pairwise constraints.

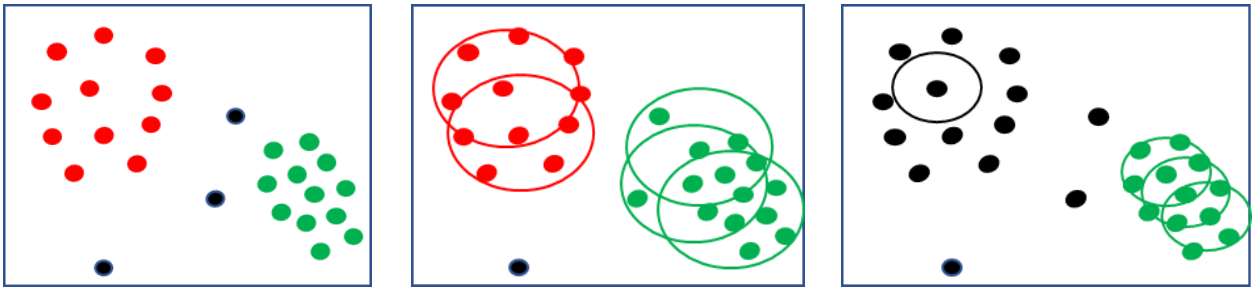


Fig. 1. Finding clusters using the DBSCAN algorithm from multi-density data (MinPt s = 5).

The paper is organized as follows. In Section II, existing literature on the given research is highlighted and critically discussed. In Section III, the research discusses the proposed algorithm and its working. In Section IV, the dataset and evaluation metrics are briefly discussed before highlighting the results of the search and its comparison to existing state-of-the-art systems. In Section V, the research is concluded along with directions for future research.

II. Related Work

In this section, existing literature on clustering machine learning algorithms is reviewed. Since the scope of this research is limited towards density-based algorithms thus it will focus on that as well as active learning methods.

In a study by Ankerst et al., an algorithm named OPTICS was proposed for the sole purpose of clustering analysis [1]. This algorithm generates two key parameters for clustering multi-density data including 1) core distance and 2) reachability. This additional information is added to the dataset to enable other algorithms to use the information for performing density-based clustering. The evaluation results showed that the algorithm was more efficient using these newly generated parameters over using existing information given in the dataset.

In another research, authors presented a clustering algorithm named DBSCAN, which is capable of finding clusters having various shapes as well as considering multiple densities [8]. However, DBSCAN algorithm faces limitation in finding clusters as they are reliant on *Eps* and *MinPts* parameters, where these values are incorrectly applied. The two studies by Jahirabadkar et al. and Liu et al. also utilize k-nearest neighbor of all selected objects so as to get the density for a given dataset. The authors further utilize various *Eps* values for sparse and dense clusters [2-3].

To address the issue of finding clusters using a dataset containing high dimensional data in distinct sizes, shapes, and density, Ertoz et al. suggested to check the points in the neighborhood of the clusters and use that information to check the similarity between the points [5]. Using this, the algorithm is able to define clusters using the defined points for those regions.

With the aim of detecting clusters that may have a unique shape or size, Liu et al. provided a variant of the DBSCAN algorithm. This algorithm as named as Entropy and Probability Distribution (EPDCA) [9]. The evaluation results conducted on benchmark datasets for the EPDCA algorithm show improved performance over other algorithms that were evaluated on the same dataset. This work was improved further by providing an optimized combinatorial clustering algorithm specifically designed for noisy performance. This algorithm is vital, particularly for data that is large with random sampling [16]. The result shows that the proposed clustering algorithm outperformance various traditional approaches that are compared with.

Kim et al. improved their work further by proposing an approximate adaptive AA-DBSCAN algorithm. This algorithm improved efficiency by reducing the time taken to calculate the parameters required to find clusters having multiple density points. The algorithm uses a density layer tree to distinguish between sparse and dense regions [10]. Kim et al. enhanced AA-DBSCAN further with *k*AA-DBSCAN that uses k^{th} nearest neighbors technique. The newer algorithm showed improvement in approximating the Eps distance but requires a longer running time [10].

In another research, authors proposed the GCMDDBSCAN algorithm, which is based on the DBSCAN algorithm. The authors proposed this algorithm to allow better performance when dealing with large databases [11]. The evaluation results showed improved performance on large databases over other algorithms evaluated on the same databases.

Constraint-based algorithms have emerged as extensions to traditional unsupervised clustering algorithms [17-19]. These methods incorporate constraints into the clustering process to enhance the learning of similarity metrics. Recently, several constraint-based clustering algorithms have been developed. For instance, Ruiz et al. introduced a pairwise-constrained clustering technique called C-DBSCAN, which leverages pairwise restrictions to enhance clustering accuracy [17]. The algorithm does this by determining the points that should be linked and the ones that should be avoided. This allows the algorithm to form clusters having unique shapes and sizes using the available constraints. In another study, a semi-supervised clustering algorithm (SSDBSCAN) is proposed that uses information provided in the dataset for evaluating density parameters [18]. The evaluation results showed better performance in comparison to other algorithms. However, SSDBSCAN can only be applied on smaller datasets.

Active learning in supervised machine learning has been a topic of interest for several decades when it comes to solving classification problems [20-22]. However, active learning for

clustering has acquired significant attention in recent years as it enhances the efficiency and accuracy of unsupervised learning algorithms [23]. Researchers have integrated active learning methods with different clustering algorithms. Active learning methods focused on uncertainty sampling, querying data points near cluster boundaries or with high uncertainty to improve cluster performance. Recently active semi-supervised clustering uses pairwise constraints to improve the clustering performance, especially for noisy, unbalanced or high-dimensional data. Thus, active learning methods can enhance clustering performance while reducing labeling costs [24-26].

III. ASS-DBSCAN ALGORITHM

This section covers ASS-DBSCAN that can rapidly and efficiently identify clusters for multi-density data having different shapes and sizes, with the knowledge of dealing with points that do not fall within the cluster. This algorithm comprises two main steps. 1) Partitioning dataset: divide the dataset into multiple density levels and generate pairwise constraints for linking and not linking points 2) Clusters are formed from the pairwise constraint information generated in step 1.

A. Generating Density levels

Let us suppose, that $X = \{x_1, x_2, \ldots, x_n\}$ represent data points where the i-th point can show nominated as $x_i = \{x_{i1}, x_{i2}, \ldots, x_{im}\}$. We start by computing the local density function for each data point. As described in Eq. (1), the local density is calculated by summing the distances between a point and its nearest neighbors. Next, the data points are ranked in descending order according to their local density values. This will allow the algorithm to calculate the difference in density between the point and its adjacent point x_i and x_{i+1} denoted by *DENVAR* (x_i, x_{i+1}) as described in Eq. (3).

$$DEN(x) = \sum_{i=1}^{k} D(x, y_i) \quad (1)$$

$$D(x, y_i) = \sqrt{\sum_{j=1}^{m} (x_j - y_j)^2} \quad (2)$$

Here $D(x, y_i)$ shows the Euclidean distance between point x and its k-nearest neighbors y_i.

$$DENVAR(x_i, x_{i+1}) = \frac{|DEN(x_{i+1}) - DEN(x_i)|}{DEN(x_i)} \quad (3)$$

After getting the density variation between data points we will notice that the variation for points that are within the proximity of the density is small and there is some distinct variation between the different densities. Thus, all density level sets (*DLS*) are collected by computing the unique density variations.

To simplify the explanation of our algorithm, we provide an example using a dataset that contains four clusters with varying densities, as explained in Fig. 2(a). We compute the density for each point of this dataset and sort them in descending order to get the results in Fig. 2(b). Fig. 2(b) contains four relatively smooth lines which accordingly represent four density levels and some sharp dips. After computing distinct variation, it can be seen that the smooth lines and sharp waves in Fig. 2(b) correspond to the density level sets and sharp change of two density levels that is highlighted in Fig. 2(c) respectively.

Each smooth line needs to be extracted to acquire the density level for that set. This can be achieved by using a density variation threshold τ. This can allow a dataset having multi-density data to be converted into a dataset containing multiple density level sets. Each Density level set should have data points with approximate densities. Points x_i and x_j are considered to be in the same density level set if the following criteria is met:

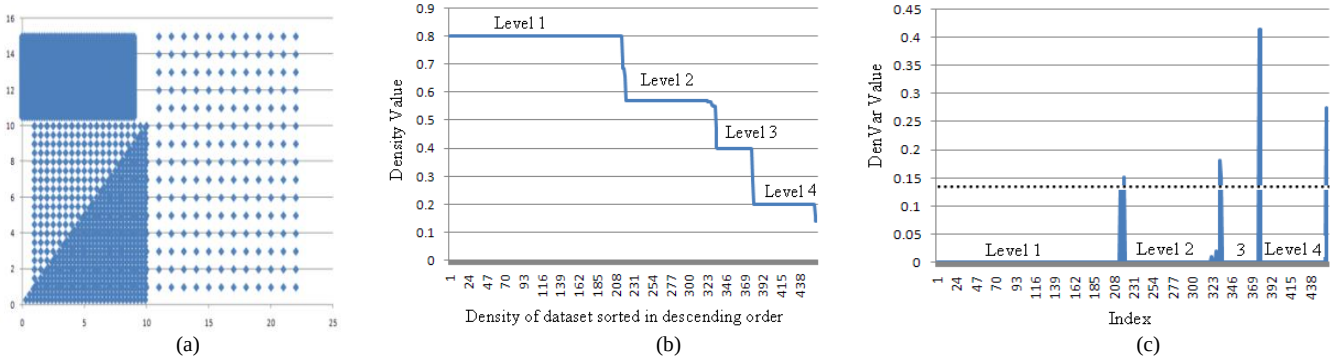
$$x_i, x_j \in DLS \quad if\ DENVAR(x_i, x_j) \leq \tau \quad (4)$$

The τ is calculated using the density variation values (*DVList*) as follows:

$$\tau = E(DVList) + \sigma(DVList)$$

where, σ is standard deviation of *DVList* and E is the mathematical expectation.

DVList values indicate that there are only a handful of points that have large DENVAR values. These points can be used to divide the dataset and transform into one having multiple density levels set.



(a) (b) (c)

Fig. 2. Density and density variation for multi-density dataset.

After splitting the dataset into various density level sets, we generate a set pairwise constraints for performing clustering. The points that must be linked should be picked from the same density level and vice-versa Thus, we need to select the most informative objects from each density level set which have the highest neighborhood density. Now, we can generate a set of not to be linked constraints between the selected objects from different density level sets. Also, generate a set of constraints

for the points that must be linked between the most informative objects and their *k*-nearest neighbors objects in the same density level set. We summarize our strategy for selecting active pairwise constraints in Algorithm 1.

Algorithm 1. Active Pairwise Constraints

1. Initialization: $N_1 = \{x\}$, where x is a random point in X; $N = N_1$; $l = 1$; $q = 0$;
2. **while** $q<Q$
3. Select the most informative point x^* with the highest neighborhood density.
4. **for** each neighborhood $N_i \in N$
5. Query x^* against any data point $x_i \in N_i$;
6. q++;
7. **if** a must-link established between x^* and x_i
8. Add x^* to neighborhood N_i;
9. **end if**
10. **end for**
11. **if** no must-link is achieved
12. Add cannot-link between x^* and all points in $\{N_i\}_{i=1}^{l}$;
13. l++; create new neighborhood $N_l = \{ x^*\}$; $N = N \cup N_l$
14. **end if**
15. **end while**
16. **return** set of active pairwise constraints

First, we select point having the most information, which has highest neighborhood density and query it against the points in each neighborhood. If a point is returned that must be linked, then this operation needs to be performed again. This process is repeated until a must does not link point is returned.

We can generate more constraints using the transitive closure between the objects in pairwise constraints. For example, if we have must-link between the two objects x_i and x_j and must-link between the two objects x_i and x_k, then we get must-link between the two objects x_j and x_k. Also, if we have must-link between the two objects x_i and x_j and cannot-link between the two objects x_i and x_k, then we get cannot-link between the two objects x_j and x_k. The partitioning dataset into different density level sets and generating the pairwise constraints are presented in Algorithm 2.

Algorithm 2: Generating density level sets with pairwise constraints

1. Calculate the density function for every point in the dataset.
2. Sort the points based on their density in descending order.
3. Determine the density variation between each pair of consecutive data points.
4. Group the dataset into different Density Level Sets (*DLS*) based on the computed density variations.
5. Create a cannot-link constraint between data points in different density level;
6. **For** each density level set (DLS_i)
7. Call Active Pairwise Constraints algorithm.

B. Expanding Clusters

When the *Eps* value is the same, clusters with different densities cannot be found. For every density level set, the *Eps* the value is approximated. The associated *Eps* for a particular density level set (*DLS*) can be boosted by setting the maximum *DEN* value. Some of the points selected may be classified as border objects or noise, which could affect the *Eps* value. Therefore, this issue is resolved by calculating Eps_i for DLS_i as follows:

$$Eps_i = \mathrm{max}DEN(DLS_i) \cdot \sqrt{\frac{medianDEN(DLS_i)}{meanDEN(DLS_i)}} \qquad (5)$$

where, meanDEN, medianDEN and maxDEN, are the mean, median and maximum density of DLS_i respectively.

Now, we explain an example to demonstrate the effect of the proposed algorithm on a dataset with different densities. First, we assume that the parameter MinPts = 5 and compute the parameter Eps for each density according to Eq. (5). Next, as shown in Fig. 3(a), we have two different densities and set of noise points. Thus, we have two values for the Eps parameter (1.7 and 0.9) for the sparse and dense density respectively as shown in Fig. 3(b) and Fig. 3(c).

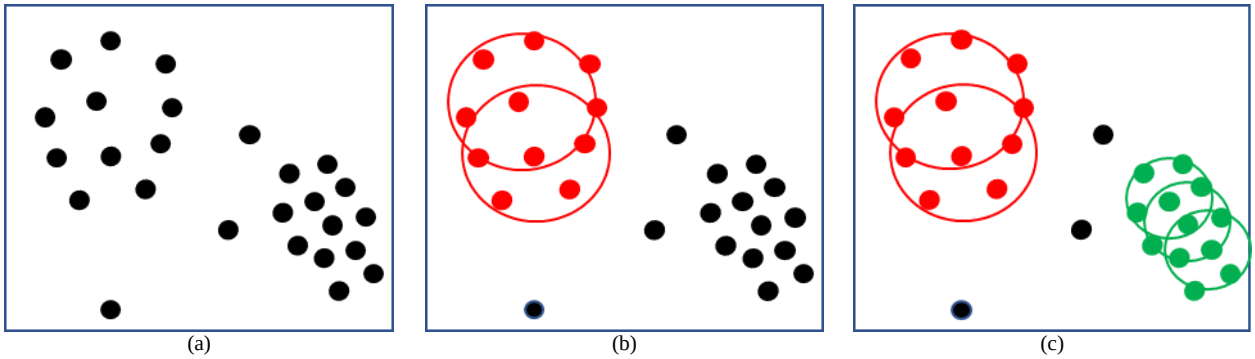
With density levels partitioning and parameters estimation done, pairwise constraints are used for each density level cluster to increase the cluster size if necessary as shown in Algorithm 3:

Algorithm 3. ASS-DBSCAN

1. Partitioning the dataset and generating pairwise constraints using algorithm 1;
2. **For** each density level set (DLS_i)
3. Estimate the parameter Eps_i ;
4. Initialize all data points in DLS_i as *UNCLUSTERED*;
5. *ClusterNum* = 0;
6. **For** each $x \in DLS_i$
7. **If** data *x not CLUSTERED* then
   Count the number of data points in *x's Eps*-neighborhood;
   a. **If** the number of data points in *neighborhood* < *MinPts*
      add x to NOISE set;
      Else
      Add x to the current cluster ClusterNum;
   b. **If** there are must-link constraints between *x* and *y*
      For each point *y* in *ML*(*x*, *y*)
      Add y to the current cluster ClusterNum;
   c. **For** each point *z* in neighborhood **do**
      If *z* is not C*LUSTERED* and does not violated cannot-link constraints then
      Add z to the current cluster ClusterNum;
8. *ClusterNum* = *ClusterNum* +1;
9. Return the set of clusters.

- In Step 7(a), after accurately estimating the *Eps* parameter, we proceed to compute the *Eps*-neighborhood for each unclustered point. We then compare the number of data points within this neighborhood to the specified *MinPts* parameter to decide regarding the classification of the data point as part of a cluster or as noise. If the count of points in the *Eps*-neighborhood is less than *MinPts*, we classify the data point as noise. Conversely, if the number of points meets or exceeds *MinPts*, we incorporate the data point into the current cluster, thereby refining the clustering process.

- In Step 7(b), we address the Must-link constraints. If a point has any Must-link constraints with other points, all of these points are assigned to the current cluster.



(a) (b) (c)

Fig. 3. Example of different *Eps* values with varying densities (MinPt s = 5).

- In Step 7(c), we consider the Cannot-link constraints. For each unclustered point within the current neighborhood that does not violate any Cannot-link constraints with the existing points in the current cluster, these points will also be included in the current cluster.

C. Time Complexity

This is divided into two parts. The first part is the time for dividing the dataset into different densities by calculating the density values. As per our calculations it leads to a time complexity of O($n$log$n$); the ordering of density values consumes O($n$log$n$); the other processes have the same runtime complexity of O(n). The second part is the time for expanding clusters, where the time required for a neighborhood query of a single object is O(n) and neighborhood query is is executed for every n points present in the dataset. The time of this part is O(m+n^2); where m is the number of pairwise constraints. However, using an index structure like R*-tree, minimize the time required for a neighborhood query of a single object to O($logn$). Thus the time required for this part is O($m + n\ log\ n$). The total runtime complexity of the proposed algorithm is O($m + n\ log\ n$).

IV. Experimental Results

This section evaluates the proposed algorithm against other baseline algorithms on a real dataset from the UCI repository. Each dataset was labeled with the number of data points, features, and cluster labels respectively as follows: Class (214,10,6), Ecoli (336,8,8), Ionosphere (351,34,2), Liver (345,6,2), Breast (683,9,2), Yeast (1484,8,10), Waveform (5000,21,3), Segment (2310,19,7) and Magic (19020,10,2). The evaluation of these algorithms is conducted using Normalized Mutual Information (*NMI*) as the clustering validation metric that is defined as follows:

$$NMI = \frac{I(X;Y)}{\big(H(X) + H(Y)\big)/2}$$

A. Clustering Performance

In this subsection, we explain the clustering performances of the algorithms C-DBSCAN, SSDBSCAN, AA-DBSCAN, and ASS-DBSCAN across the nine real-world datasets mentioned earlier. Table I presents the parameters values of each algorithm. As the parameter settings will be different according to each dataset, we employ a range of values for each parameter as detailed in Table I.

TABLE I. Parameters Values of Each Algorithm

Algorithm	Parameters
C-DBSCAN	*Eps* ∈ [0.1, 0.5] and *MinPts* ∈ {3, 4, . . . , 10}
SSDBSCAN	*Eps* ∈ [0.1, 0.5] and *MinPts* ∈ {3, 4, . . . , 10}
AA-DBSCAN	*MinPts* ∈ {3, 4, . . . , 10}
ASS-DBSCAN	*MinPts* ∈ {3, 4, . . . , 10}

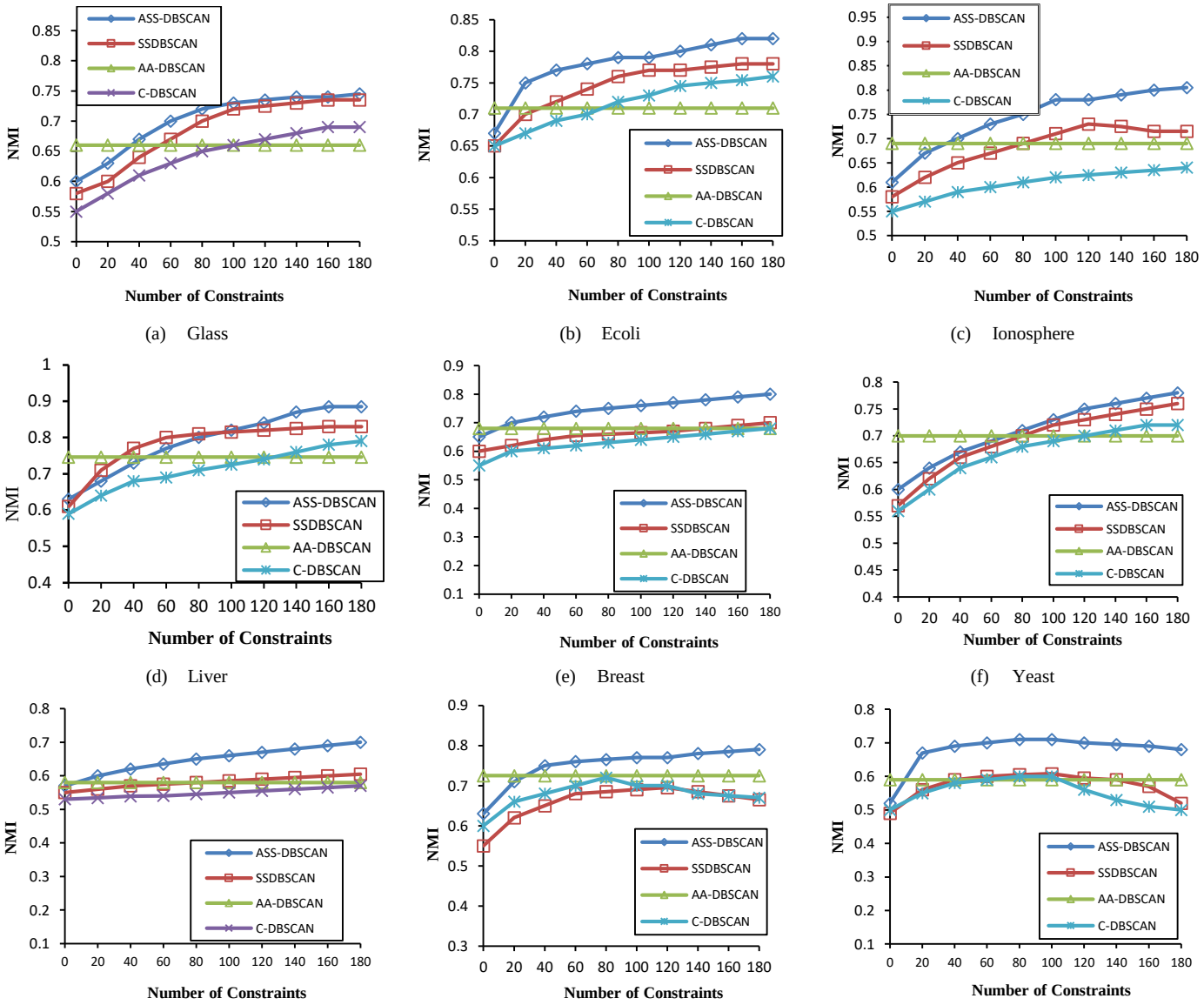
We present the clustering performance result based on NMI of ASS-DBSCAN in comparison to compared algorithms with varied constraints on the real datasets. Each algorithm is shown with its own unique color in Fig. 4. For each dataset, a number of constraints were utilized with respect to the NMI. The result demonstrates that ASS-DBSCAN performance better than the compared algorithms in general. For example, as illustrated in Fig. 4(a), when utilizing 20 constraints, ASS-DBSCAN achieves a performance level exceeding 0.6 NMI, and with 80 constraints, ASS-DBSCAN achieves NMI of more than 0.7, respectively.

It can be observed that SSDBSCAN comes second in terms of effectiveness, followed by C-DBSCAN. Even though AA-DBSCAN has a higher performance in the early stages. However, with the increased number of pairwise constraints, AA-DBSCAN has consistently shown in all the datasets its static performance with stable and non-improved result. This can be explained since algorithms are an unsupervised clustering algorithm. Hence, it is not effective with pairwise constraints. Also, we can observe that the performance of ASS-DBSCAN can slowly increases with increased number of pairwise constraints as shown for Glass and Breast datasets. However, its performance surpasses the other algorithms.

B. Running Time Evaluation

This section discusses the running time of the ASS-DBSCAN and other algorithms that have been evaluated. Fig. 5 shows the execution times on the six datasets (containing pairwise constraints) used for evaluating. Lower execution time is considered better and vice versa. From Fig. 5, SSDBSCAN achieved the lowest performance of all pairwise constraints. AA-DBSCAN also has static value as shown in Fig. 5. These static values manifest as the number of

constraints are increased. We observed that the ASS-DBSCAN on most datasets achieve relatively high execution time with low constraints, however, overall, the execution time drops, and better performance was recorded for ASS-DBSCAN. Looking at all the experiments conducted on the other datasets, we can see that consistently ASS-DBSCAN performs better by achieving lower execution times in comparison to other algorithms.

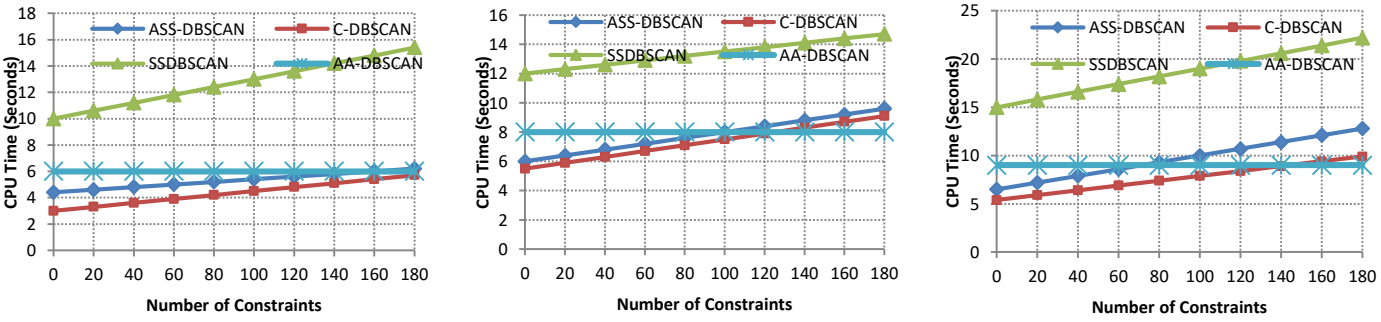


(a) Glass (b) Ecoli (c) Ionosphere

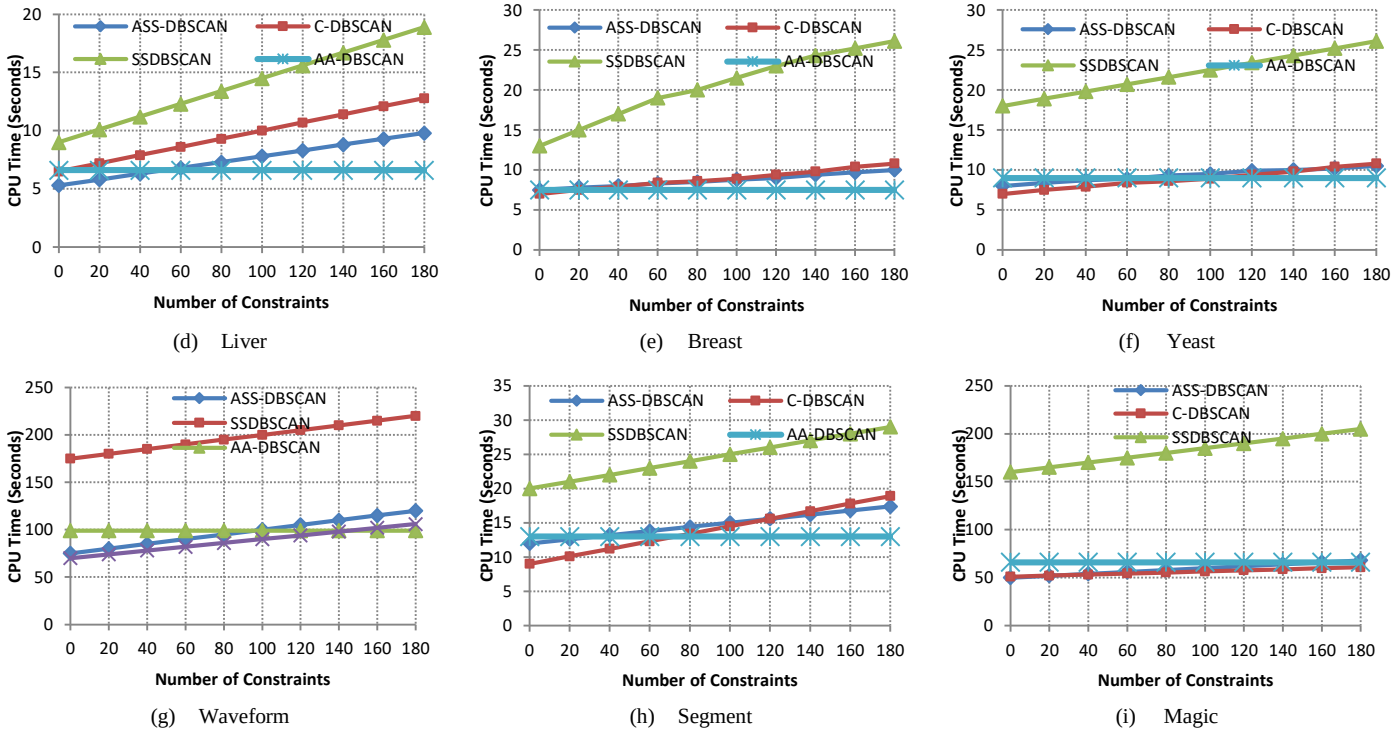
(d) Liver (e) Breast (f) Yeast

(g) Waveform (h) Segment (i) Magic

Fig. 4. Clustering performance based on NMI with different numbers of pairwise constraints.



(a) Glass (b) Ecoli (c) Ionosphere



(d) Liver (e) Breast (f) Yeast

(g) Waveform (h) Segment (i) Magic

Fig. 5. Execution time with different numbers of pairwise constraints.

V. Conclusions and Future Work

In this paper, we propose ASS-DBSCAN for clustering different density data with a set of active pairwise constrints. By examining the statistical properties of the dataset's density variation, the suggested algorithm divides it into multiple density level sets, which are subsequently expanded using active pairwise constraints. The algorithm was evaluated for performance and execution time on the real datasets again other algorithms. The evaluations results showed that the algorithm not only performed better in achieving its goals but also took less time in order to do so. Future work includes the extension the current work to cluster more complex data from real-life streaming applications. In addition, we aim to develop a comprehensive system that integrates active learning and semi-supervised learning techniques to be applied on different applications.

Acknowledgment

This work was funded by the University of Jeddah, Saudi Arabia, under grant No. (UJ-21-DR-134). The authors, therefore, acknowledge with thanks the university's technical and financial support.

References

[1] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "OPTICS: ordering points to identify the clustering structure," *ACM Sigmod Rec.*, vol. 28, no. 2, pp. 49–60, 1999.

[2] P. Liu, D. Zhou, and N. Wu, "VDBSCAN: varied density based spatial clustering of applications with noise," in *2007 International conference on service systems and service management*, 2007, pp. 1–4.

[3] X. Zhang, and Z. Shibo "WOA-DBSCAN: Application of Whale Optimization Algorithm in DBSCAN Parameter Adaption," IEEE Access, 2023.

[4] A. A. Almazroi, and W. Atwa, "An improved clustering algorithm for multi-density data," Axioms vol 11, no 8, 2022.

[5] L. Ertöz, M. Steinbach, and V. Kumar, "Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data," in *Proceedings of the 2003 SIAM international conference on data mining*, 2003, pp. 47–58.

[6] A. Fahim, "Adaptive Density-Based Spatial Clustering of Applications with Noise (ADBSCAN) for Clusters of Different Densities," Computers, Materials & Continua vol 75, no 2, 2023.

[7] W. Atwa, and K. Li. "Constraint-based clustering algorithm for multi-density data and arbitrary shapes." *In Industrial Conference on Data Mining*, pp. 78-92. Springer, Cham, 2017.

[8] M. Ester, H.P. Kriegel, J. Sander, and X. Xu, 'A density based algorithm for discovering clusters in large spatial databases with noise'. In *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining*, pp. 226–231, 1996.

[9] X. Liu, Q. Yang, and L. He, "A novel DBSCAN with entropy and probability for mixed data," *Cluster Comput.*, vol. 20, no. 2, pp. 1313–1323, 2017.

[10] J.-H. Kim, J.-H. Choi, K.-H. Yoo, and A. Nasridinov, "AA-DBSCAN: an approximate adaptive DBSCAN for finding clusters with varying densities," *J. Supercomput.*, vol. 75, no. 1, pp. 142–169, 2019.

[11] L. Zhang, Z. Xu, and F. Si, "GCMDDBSCAN: multi-density DBSCAN based on grid and contribution," in *2013 IEEE 11th International Conference on Dependable, Autonomic and Secure Computing*, 2013, pp. 502–507.

[12] M. A. Masud, J. Z. Huang, M. Zhong, and X. Fu, "Generate pairwise constraints from unlabeled data for semi-supervised clustering," *Data Knowl. Eng.*, vol. 123, p. 101715, 2019.

[13] W. Atwa, and K. Li. "Active query selection for constraint-based clustering algorithms." In *International Conference on Database and Expert Systems Applications*, pp. 438-445. Springer, Cham, 2014.

[14] W. Atwa, and M. Emam. "Improving Semi-Supervised Clustering Algorithms with Active Query Selection." *Advances in Systems Science and Applications* 19, no. 4 (2019): 25-44.

[15] W. Atwa, "A Supervised Feature Selection Method with Active Pairwise Constraints." In *Proceedings of the 11th International Conference on Informatics & Systems (INFOS 2018)*. 2018.

[16] J. Kim, W. Lee, J. J. Song, and S.-B. Lee, "Optimized combinatorial clustering for stochastic processes," *Cluster Comput.*, vol. 20, no. 2, pp. 1135–1148, 2017.

[17] C. Ruiz, M. Spiliopoulou and E. Menasalvas. C-DBSCAN: Density-Based Clustering with Constraints. In Proceedings of the International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, pages 216- 223, 2007.

[18] L. Lelis and J. Sander. Semi-Supervised Density-Based Clustering. In Proceeding of the Ninth IEEE International Conference on Data Mining, 842-847, 2009.

[19] W. Atwa, and K. Li. "Semi-supervised Clustering Method for Multi-density Data." *In International Conference on Database Systems for Advanced Applications*, pp. 313-319. Springer, Cham, 2015.

[20] P. Zhou, S. Bicheng, L. Xinwang, D. Liang, and L. Xuejun, "Active Clustering Ensemble With Self-Paced Learning," IEEE Transactions on Neural Networks and Learning Systems, 2023.

[21] P. Kumar, and G. Atul, "Active learning query strategies for classification, regression, and clustering: a survey," Journal of Computer Science and Technology, vol 35, pp 913-945, 2020.

[22] A. Hussein, A. A. Almazroi, "Multiclass Classification for Cyber Threats Detection on Twitter", CMC-COMPUTERS MATERIALS & CONTINUA, 77(3), 2023.

[23] W. Yu, L. Xing, F. Nie, and X. Li, "An efficient semi-supervised balanced cut with hard pairwise constraints and partial labels," Knowledge-Based Systems, vol 276, 2023.

[24] J. Cai, J. Hao, H. Yang, X. Zhao, and Y. Yang, "A review on semi-supervised clustering," Information Sciences, 2023.

[25] W. Atwa, A. A. Almazroi, "Active selection constraints for semi-supervised clustering algorithms", . Int. J. Inf. Technol. Comput. Sci, 2020.

[26] A. A. Almazroi, and W. Atwa, "Semi-Supervised Clustering Algorithms Through Active Constraints", *International Journal of Advanced Computer Science & Applications*, *15*(7), 2024.

MSMA: Merged Slime Mould Algorithm for Solving Engineering Design Problems

Khaled Mohammad Alhashash[1], Hussein Samma[2], Shahrel Azmin Suandi[3*]
Intelligent Biometric Group, School of Electrical and Electronic Engineering, USM Engineering Campus, Universiti Sains Malaysia, Nibong Tebal 14300, Penang, Malaysia[1, 3]
SDAIA-KFUPM Joint Research Center for Artificial Intelligence (JRCAI), King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia[2]

***Abstract*—The Slime Mould Algorithm (SMA) has effectively solved various real-world problems such as image segmentation, solar photovoltaic cell parameter estimation, and economic emission dispatch. However, SMA and its variants still face limitations when dealing with low-dimensional optimization problems, including slow convergence and local optima traps. This study aims to develop an optimized algorithm, the Merged Slime Mould Algorithm (MSMA), to overcome these limitations and improve performance in low-dimensional optimization tasks. Additionally, MSMA introduces a novel approach by merging the Adaptive Opposition Slime Mould Algorithm (AOSMA) and the Smart Switching Slime Mould Algorithm (S2SMA), simplifying the hybridization process and enhancing optimization performance. MSMA eliminates the need for multiple initializations, avoids memory-switching requirements, and employs adaptive and smart switching rules to harness the strengths of both algorithms. The performance of MSMA is evaluated using the CEC 2005 benchmark and ten real-world applications. The Wilcoxon rank-sum test verifies the effectiveness of the proposed approach, with results compared to various SMA variations and related optimization methods. Numerical findings demonstrate superior fitness values achieved by the proposed strategy, while statistical results indicate MSMA's outperformance with a rapid convergence curve.**

Keywords*—*Slime mould algorithm; engineering design problems; metaheuristic; optimization

I. Introduction

Metaheuristic algorithms (MAs) offer valuable tools for solving complex engineering problems in a reasonable time [1]. These algorithms provide a flexible and efficient approach to optimization, enabling engineers to find near-optimal solutions in diverse domains. MAs have two main elements: exploration and exploitation abilities [2]. Exploration capability is the ability to converge to a possible global optimum with increasing solution space and randomness. On the other hand, exploitation capability refers to the ability to search more precisely in the region that the algorithm's exploration phase has identified. There are two categories of metaheuristics: population-based and single-solution-based metaheuristics [3]. Population-based approaches involve utilizing a collection of solutions, referred to as a population, to generate and substitute candidate solutions throughout the optimization procedure. Some of the popular population-based metaheuristic approaches are Particle Swarm Optimization (PSO) [4], whale optimization algorithm (WOA) [5], and Harris Hawk Optimizer (HHO) [6]. In contrast, metaheuristics that rely on a single solution-based approach involve generating a set of potential solutions derived from the current solution. Subsequently, the current solution is substituted with one of the candidate solutions during each iteration. This category involves the local search (LS)[7], Tabu search (TS)[8], and simulated annealing (SA) [9].

Single-based and population-based algorithms have benefits and are widely utilized to address various issues. Nevertheless, no single approach can solve all optimization problems [10]. Developing an optimization algorithm to address these issues is necessary, but researchers have found it challenging to design new optimization algorithms from scratch. In this direction, hybridizing meta-heuristic algorithms is the most common and successful technique. For example, on hybridizing meta-heuristic algorithms [11]–[13].

In the literature, several optimizers have emerged recently, such as SMA [14], Fitness Dependent Optimizer (FDO) [15], Black Widow Optimization Algorithm (BWO) [16], and Reptile Search Algorithm (RSA) [17]. SMA has captured considerable attention due to its smooth structure, limited parameter requirements, robustness, and flexibility in implementation. It presents itself as a valuable and efficient approach for addressing a wide range of real-world optimization problems [18], such as image segmentation [19], estimation of solar photovoltaic cell parameters [20], and economic emission dispatch [21]. Nevertheless, similar to other metaheuristic algorithms, SMA encounters challenges related to local optimality and premature convergence in some optimization problems [22], [23]. Moreover, Utilizing two random search agents from the entire population to determine the future displacement and direction based on the best search agents restricts SMA's exploitation and exploration capabilities [24]. Researchers suggested hybridized and modified variants of SMA to address these limitations.

This research article presents the hybridization of two variants of SMA, namely S^2SMA [25] and AOSMA [24]. The integration involves incorporating a set of vertical smart switching rules to govern the transition process between AOSMA and S2SM. The two algorithms were combined intelligently, where the invocation procedure exclusively occurs during the update of slime locations. This merger is unique and distinct from SAM's previous integration due to the following three advantages: no necessity for multiple initializations for different algorithms, no memory-switching needs, and

employing adaptive and intelligent switching rules to leverage the strengths of both algorithms. The main contributions of this work are outlined as follows:

- MSMA introduces a novel optimization approach by intelligently merging AOSMA and S^2SMA, setting it apart from previous SMA integrations through its streamlined operational framework, which simplifies the algorithm hybridization process.
- The MSMA eliminates the necessity for multiple initializations and memory-switching, significantly enhancing computational efficiency. This innovation reduces the algorithm's complexity and resource consumption, facilitating a more seamless optimization experience.
- Incorporating Vertical Smart Switching Rules (VSSR) enables MSMA to facilitate dynamic algorithmic switches based on problem-specific attributes, amplifying adaptability and operational efficiency. VSSR represents a critical innovation, ensuring effective navigation through complex problem spaces and significantly improving optimization performance.
- The MSMA has been rigorously validated through extensive experiments and numerical studies, demonstrating its superiority in solving optimization problems.

This paper's remaining sections are organized as follows. The pertinent studies on SMA and engineering design problems are summarized in Section II. Section III illustrates the slime mould algorithm and the proposed work in detail. Section IV presents the numerical experiment and statistical analysis. This paper's conclusion is given in Section V.

II. Related Work

As mentioned previously, SMA can be categorized into hybridized and modified forms of SMA. Many studies have investigated the idea of hybridizing SMA with other metaheuristic algorithms [26]–[30]. Among these advancements, Naik et al. [26] introduced the Equilibrium Slime Mould Algorithm (ESMA), merging the Slime Mould Algorithm (SMA) with the Equilibrium Optimizer (EO) for enhanced multilevel thresholding in breast thermogram images. ESMA aims to reduce entropic dependencies between image classes, showing improved exploration capability and efficient analysis over other optimization methods. Although it outperforms in breast thermogram analysis, suggesting potential benefits for medical diagnostics, ESMA faces challenges in specific clinical contexts and broader medical imaging applications. Further contributing to the field, Chen et al. [27] introduced CHDESMA, an improved Slime Mould Algorithm (SMA) using chaotic maps and Differential Evolution (DE). CHDESMA mitigates SMA's local optima and population diversity issues by integrating chaotic maps for initialization and DE strategies for enhanced search. Evaluations against benchmarks and real-world problems show CHDESMA's competitive performance against advanced algorithms and DE variants, emphasizing its effectiveness and contributions in diverse scenarios. Moreover, Bhandakkar and Mathew [28] proposed using Integrated Slime Mould Algorithm (ISMA) for optimal placement of a Hybrid Power Flow Controller (HPFC). ISMA combines the Slime Mould Algorithm (SMA) with WOA for enhanced searching behavior. This optimization aims to minimize system power loss and generation cost by determining optimal locations for Unified Power Flow Controllers (UPFCs) and their capacities while considering system stability constraints. Chen et al. [29] presented RCLSMAOA, merging SMA and AOA to improve optimization. Through extensive testing, it effectively combines global exploration and local exploitation strategies. Despite the success, challenges persist in high-dimensional spaces and convergence accuracy. Future work aims to refine RCLSMAOA's performance in practical engineering problems and high dimensions, potentially exploring a binary version of the algorithm for further enhancement. Finally, Ewees et al. [30] introduced GBOSMA, a hybrid method merging Gradient-Based Optimizer (GBO) and Slime Mould Algorithm (SMA) to improve global optimization and feature selection. GBOSMA enhances exploration by using SMA as a local search within GBO, achieving better performance than standard GBO, SMA, and recent algorithms in both speed and accuracy across diverse benchmarks. The results showcase GBOSMA's superiority, achieving top fitness values in 66% of global optimization functions and the highest accuracy in 93% of feature selection benchmarks. This approach holds potential for various applications like medical imaging, object detection, and weather prediction tasks.

In many investigations, modified SMA methods were presented [24], [25], [31]–[34]. The Adaptive Opposition Slime Mould Algorithm (AOSMA), as introduced by Naik et al. [24], represents an advancement in the Slime Mould Algorithm (SMA) through the integration of adaptive opposition-based learning. This enhancement significantly boosts the algorithm's exploration and exploitation capabilities, making it a powerful tool for solving complex problems. However, AOSMA is not without its limitations. It shows a marked reliance on specific problem types, indicating that its effectiveness may be constrained to particular domains. Additionally, there is a noted requirement for further validation to confirm its broader applicability across a wider range of problem scenarios. Alhashash et al. [25] introduced an enhanced optimizer named Smart Switching Slime Mould Algorithm (S^2SMA) that enhances the accuracy of face sketch recognition by fine-tuning pre-trained deep learning models, which is challenging due to limited sketch datasets. S^2SMA simultaneously fine-tunes multiple deep learning models and uses embedded rules and search operations for adaptive switching between search operations during execution. The proposed algorithm was evaluated on CEC's 2010 large-scale benchmark and two face sketch databases and outperformed other optimization techniques with a faster convergence rate. The outcomes revealed the superiority of S^2SMA in the majority of experiments. Ewees et al. [31] presented a modified version of the slime mould algorithm (SMA) called SMAMPA, which incorporates the Marine Predators Algorithm (MPA) operators as a local search strategy. The proposed feature selection technique was evaluated on twenty UCI datasets and compared with other state-of-the-art FS methods, showing superior performance in terms of efficiency and performance metrics. The SMAMPA method was also applied to real-world problems,

such as QSAR modeling and chemometrics, with promising results. Future work includes investigating SMAMPA in more complicated problems, such as multi-optimization problems and big data mining. Abid et al. [32] proposed an enhanced slime mould optimization algorithm (ESMOA) to optimize tuning parameters for a cascaded proportional derivative-proportional integral (PD-PI) controller in order to solve frequency stability problems (FSP) in multi-area power systems (MAPSs) with two-area non-reheat thermal systems. ESMOA surpassed current PID and PI controllers. Cascaded PD-PI controller designs are more reliable than GSO and CO algorithms due to ESMOA's chaotic dynamic and elite group. In time domain simulations, ESMOA beat both GSO and CO. Deng and Liu [33] proposed AGSMA, an improved variant of the slime mould algorithm, to address limitations such as insufficient exploration, slow convergence, and an imbalance between diversity and convergence. AGSMA achieved a balance between convergence and diversity through adaptive grouping, a new search mechanism, and an efficient learning operator. Experiments demonstrated that it outperformed other methods and is able to solve complex nonlinear problems. However, premature convergence in some multimodal problems needs additional study. Sharma et al. [34] presented modifications to the Slime Mould Algorithm (SMA) to make it more effective for engineering design tasks, including opposition theory and a sine cosine-based position update mechanism. These modifications were found to significantly enhance the performance of SMA on standard benchmark functions and make it suitable for demand-side management tasks.

In the evolving field of metaheuristic algorithms, recent studies have made significant strides in addressing complex engineering design problems. Samma et al. [13] pioneered the Q-learning-based Simulated Annealing (QLSA) algorithm, setting a precedent for dynamic parameter control and adaptability, albeit with scalability and exploration scope limitations. Building on this, Nadimi-Shahraki et al. [1] introduced the Gaze Cues Learning-based Grey Wolf Optimizer (GGWO), which incorporated novel search strategies inspired by wolf behavior, showing promise despite challenges in selective pressure optimization. Further contributions, such as Wang et al. [35]'s Artificial Rabbits Optimization (ARO) and Yildiz et al. [36]'s Elite Opposition-Based Learning Grasshopper Optimization (EOBL-GOA), demonstrated the algorithms' strengths in diverse engineering problems but also highlighted the need for domain-specific adaptability. Zhang et al. [37] and Yıldız et al. [38] proposed enhancements to the Slime Mould Algorithm (SMA) and introduced the Chaotic Lévy flight distribution (CLFD) algorithm, respectively, achieving improved solution quality and exploration-exploitation balance. Recent developments saw Yang et al. [39] focus on the ARSCA algorithm, addressing computational complexity while improving convergence accuracy. Abdel-Basset et al. [40] applied the Nutcracker Optimization Algorithm (NOA) to engineering problems, demonstrating the ease of implementation and high convergence speed but facing challenges in exploration-exploitation balance. Gharehchopogh et al. [41] introduced the Chaotic Quasi-oppositional Farmland Fertility Algorithm (CQFFA), which enhanced exploration and convergence via chaotic maps and the Quasi-Oppositional Binary Leader strategy, albeit with hybridization challenges. Deng and Liu [42] showcased the Multi-strategy Improved Slime Mould Algorithm (MSMA), signaling a need for enhancements in multi-objective optimization and broader domain adaptability.

Despite significant advancements in developing SMA variants, current methods still face challenges in broader applicability and often struggle with slow convergence and local optima traps in low-dimensional optimization problems. This gap highlights the need for improved solutions that can overcome these limitations. The proposed approach addresses these challenges by integrating SMA variants with adaptive mechanisms, enhancing computational efficiency, reducing the reliance on multiple initializations, and simplifying the hybridization process, offering a more robust and effective solution for complex optimization tasks.

III. Proposed SMA-Based Method

A. The original Slime Mould Algorithm (SMA)

Li et al. [14] introduced the SMA as an innovative optimization mechanism for global optimization. SMA focuses on the behavior and morphological changes that the slime mould Physarum polycephalum undergoes during nutrient acquisition. Approaching, wrapping, and grabbing food are the three stages of SMA.

1) Approaching food: The concentration of odor in the air is essential for a slime mould to approach food. This contraction pattern when nearing food is defined by Eq. (1):

$$\overrightarrow{X(t+1)} = \begin{cases} \overrightarrow{X_b(t)} + \overrightarrow{vb} \cdot \left(\vec{W} \cdot \overrightarrow{X_A(t)} - \overrightarrow{X_B(t)}\right), r_2 < p \\ \overrightarrow{vc} \cdot \overrightarrow{X(t)}, r_2 \geq p \end{cases} \quad (1)$$

where the parameter $\overrightarrow{vb}$ takes values within the range of [-a, a], while $\overrightarrow{vc}$ gradually decreases from one to zero in a linearly. The position $\overrightarrow{X_b}$ refers to the current location of an individual with the highest concentration of odor detected. $\vec{X}$ represents the current location of slime mould. $\overrightarrow{X_A}$ and $\overrightarrow{X_B}$ denote two randomly selected individuals from a population of size n. The variables t and r_2 indicate the current iteration number and a random value between 0 and 1, respectively. The weight of slime mould is represented by $\vec{W}$. The parameter p is computed using Eq. (2):

$$p = tanh|S(i) - DF| \quad (2)$$

where $i \in 1,2, \ldots, n$. The fitness of the current location $\vec{X}$ is denoted by $S(i)$, while DF denotes the best fitness achieved across all iterations. The formula for computing $\overrightarrow{vb}$ can be found in Eq. (3), and the value of a is provided in Eq. (4).

$$\overrightarrow{vb} = [-a, a]) \quad (3)$$

$$a = arctanh(-\left(\frac{t}{max\_t}\right) + 1) \quad (4)$$

Here, max_t refers to the maximum number of iterations. The formula for calculating $\vec{W}$ is presented in Eq. (5), while its $SmellIndex$ is defined in Eq. (6).

$$\overrightarrow{W(SmellIndex(\iota))} = \begin{cases} 1 + r_3 \cdot log\left(\frac{bF - S(i)}{bF - wF} + 1\right), condition \\ 1 - r_3 \cdot log\left(\frac{bF - S(i)}{bF - wF} + 1\right), \ others \end{cases} \quad (5)$$

$$SmellIndex = sort(S) \quad (6)$$

where, condition represents that S(i) must be ranked within the top fifty percent of the entire population. The variable r_3 denotes a random value ranging from 0 to 1. The best fitness value achieved during the current iteration process is represented as bF, while the worst fitness value is denoted as wF. SmellIndex corresponds to the sequence of fitness values arranged in ascending order.

2) Wrapping food: Updates to the position of slime mould can be calculated using the formula given in Eq. (7):

$$\overrightarrow{X^*(t+1)} = \begin{cases} r_4 \cdot (UB - LB) + LB, r_1 < z \\ \overrightarrow{X_b(t)} + \overrightarrow{vb} \cdot \left(W \cdot \overrightarrow{X_A(t)} - \overrightarrow{X_B(t)}\right), r_2 < p \ \ and \ r_1 \geq z \\ \overrightarrow{vc} \cdot \overrightarrow{X(t)}, r_2 \geq p \ \ and \ r_1 \geq z \end{cases} \quad (7)$$

where, r_1, r_2 and r_4 are randomly selected from the interval [0,1]. B and UB represent the lower and upper bounds of the search range, respectively. The p value signifies the probability associated with the presence of slime mould, while z is a parameter with a constant value of 0.03.

3) Grabbling food: To represent the slime mould venous width changes, SMA utilizes the vectors $\overrightarrow{W}$, $\overrightarrow{vb}$, and $\overrightarrow{vc}$. $\overrightarrow{W}$ reflects the oscillating frequency of slime mould, which is determined by analyzing the quality of the food source. It helps update the speed of movement towards the food source, aiding the slime mould in selecting the most suitable food source.

The values of $\overrightarrow{vb}$ and $\overrightarrow{vc}$ undergo random oscillations within specific ranges. The vector $\overrightarrow{vb}$ ranges from -a to a, while $\overrightarrow{vc}$ ranges from -1 to 1. As the iterative process progresses, these vectors converge towards zero.

The variation in $\overrightarrow{vb}$ replicates the slime mould's behaviour when it encounters a new food source. Even if an improved food supply has been identified, the slime mould continues to explore other locations by separating some organic matter. This behaviour increases the chances of finding higher-quality food sources and improves the optimization of local problems. For further details on the SMA, refer to the study conducted by Li et al. [14].

B. The Proposed Merged Slime Mould Algorithm

MSMA is a novel optimization method that combines two variants SMA: AOSMA and S^{2}SMA. This merger distinguishes itself from previous integrations by providing three primary advantages: the elimination of the necessity for multiple initializations for different algorithms, avoidance of memory-switching requirements, and the incorporation of adaptive and intelligent switching rules, known as the Vertical Smart Switching Rules (VSSR).

The formulation of VSSR involves incorporating four embedded vertical smart switching rules to control the recall ratio between AOSMA and S^{2}SMA during slime position updates. The activation of VSSR is dependent on the occurrence of specific events, comprising seven parameters: $AOSMA\_EN$ parameter, S^2SMA_EN parameter, $EVAL\_C$ parameter, $AOSMA\_C$ parameter, S^2SMA_C parameter, $VER\_SUCCESS_{LEADER}$ parameter, and per. The first parameter is $AOSMA\_EN$. It will take either zero or one. If AOSMA is chosen to update slime positions, it will be one; otherwise, it will be zero. The second parameter is S^2SMA_EN. It will take either zero or one. If S^{2}SMA is chosen to update slime positions, it will be one; otherwise, it will be zero. The third parameter, $EVAL\_C$, evaluation counter represents the number of iterations required to evaluate the performance of the two algorithms and is computed using Eq. (8). The fourth and fifth parameters, $AOSMA\_C$ and S^2SMA_C, respectively, count the number of times each algorithm successfully finds a new leader within $EVAL\_P$ iterations when used to update slime positions. The sixth parameter, $VER\_SUCCESS_{LEADER}$, assumes one value if any methods can find a new leader and zero otherwise. The final parameter, per, takes on a value within the range of [0,1], and its value depends on the rules to be applied, which will be further expounded in the ensuing section.

$$EVAL\_P = \epsilon * max\_t \quad (8)$$

where, ϵ is a constant parameter of 0.02, its value is affordable, which was selected to ensure timely switching. However, increasing this value would result in slower switching, perhaps introducing bias. Conversely, decreasing the value would lead to faster switching, hence increasing complexity. Moreover, this parameter is adjustable based on the nature of a given problem.

The first and second rules are depicted in Fig. 1 and Fig. 2 respectively. They were developed to update the value of $AOSMA\_C$ and S^2SMA_C, which indicates the number of successes for each approach during the process of finding a new leader. Both rules will be checked in every iteration. The first rule will apply if AOSMA is called while updating the slime position and a new leader is found. Thus, $AOSMA\_C$ will be updated. The second rule will apply if S^{2}SMA is called while updating the slime position and a new leader is found. Thus, S^2SMA_C will be updated. It should be noted that the proposed method will give AOSMA and S^{2}SMA equal priority to change slime positions during the first $EVAL\_P$. During the process, the values of both $AOSMA\_C$ and S^2SMA_C will be updated as explained in Rule1 and Rule2.

RULE 1: IF $AOSMA\_EN == 1$ **AND** $VER\_SUCCESS_{LEADER} == 1$
THEN $AOSMA\_C = AOSMA\_C + 1$

Fig. 1. RULE 1 To count the number of times AOSMA was successful during a given period.

RULE 2: IF $S^2SMA\_EN == 1$ **AND** $VER\_SUCCESS_{LEADER} == 1$
THEN $S^2SMA\_C = S^2SMA\_C + 1$

Fig. 2. RULE 2 To count the number of times S^{2}SMA was successful during a given period.

The third and fourth rules are shown in Fig. 3 and Fig. 4, respectively. They are formulated to calculate per, which is the

ratio of AOSMA and S^{2}SMA calling to update slime positions during the subsequent $EVAL\_P$. This value depends on the values of $AOSMA\_C$ and S^2SMA_C, as explained in Rule1 and Rule2. If AOSMA and S^{2}SMA cannot find a new leader during the current $EVAL\_P$, both methods will be given an equal chance over the subsequent $EVAL\_P$; otherwise, the third and fourth rules will be applied. The third rule is applied if the value of $AOSMA\_C$ is greater than S^2SMA_C; otherwise, the fourth rule will be applied.

RULE 3: IF $AOSMA\_C > S^2SMA\_C$

THEN $per = AOSMA\_C/(AOSMA\_C + S^2SMA\_C)$

Fig. 3. RULE 3 to compute the probability of AOSMA being called within the specified period.

RULE 4: IF $AOSMA\_C \leq S^2SMA\_C$

THEN $per = 1 - (S^2SMA/(AOSMA\_C + S^2SMA\_C))$

Fig. 4. RULE 4 to compute the probability of S^{2}SMA being called within the specified period.

In the SMA algorithm, the arctanh function is utilized to calculate the value of parameter a in Eq. (4). However, it has been observed that the arctanh function can lead to programming warnings/errors [43]-[45]. To enhance the performance and stability of the standard SMA algorithm and achieve faster convergence with reduced warnings/errors during program execution, alternative controlling equations, such as the cosine function, have been proposed as viable solutions [44]. In this study, the value of parameter a was computed using Eq. (9), which was directly obtained from [45].

$$a = 1 + cos\left(\frac{t}{max\_t} \cdot \pi\right) \quad (9)$$

where max_t is the maximum number of iterations and t is the current iteration.

Fig. 5 depicts the complete stages of the proposed MSMA algorithm.

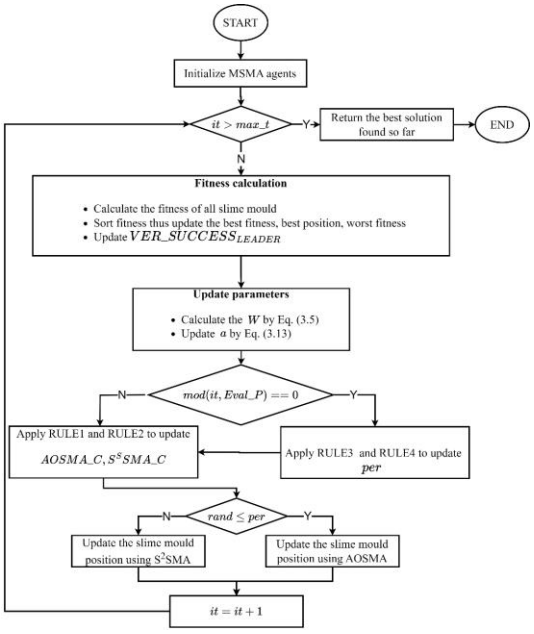


Fig. 5. Flow chart of the proposed MSMA algorithm.

IV. Experimental Results and Analysis

In this analysis section, several experiments were conducted to demonstrate MSMA's efficacy. Three case studies were investigated, comprising basic benchmark problems CEC 2005 [46] and seven engineering designs.

A. Evaluation on Basic Benchmark Functions

In this section, a total of 23 CEC 2005 [46] continuous benchmarks were used, categorized into seven unimodal (F1-F7), six multimodal (F8-F13), and ten fixed-dimensional multimodal functions (F14-F23), as illustrated in Table I, Table II, and Table III. Unimodal functions assess exploitation efficiency with one global optimum, while multimodal functions (F8-F13) evaluate exploration and local optima avoidance. Fixed-dimensional tests (F14-F23) provide a middle ground with fewer local optima, gauging the algorithm's balance between exploitation and exploration.

TABLE I. Description of Unimodal Benchmark Functions

Function	Description	Dim	Range	f_{min}
$F_1(X) = \sum_{j=1}^{D} {x_j}^2$	Sphere	30	[-100,100]	0
$F_2(X) = \sum_{j=0}^{D} \lvert x_j \rvert + \prod_{j=0}^{D} \lvert x_j \rvert$	Schwefel 2.22	30	[-10,10]	0
$F_3(X) = \sum_{j=1}^{D} (\sum_{k=1}^{j} x_k)^2$	Schwefel 1.2	30	[-100,100]	0
$F_4(X) = max_j\{\lvert x_j \rvert, 1 \leq j \leq D\}$	Schwefel 2.21	30	[-100,100]	0
$F_5(X) = \sum_{j=1}^{D-1} [\ 100(x_{j+1} - {x_j}^2)^2 + (x_j - 1)^2]$	Rosenbrock	30	[-30,30]	0
$F_6(X) = \sum_{j=1}^{D} ([x_j + 0.5])^2$	Step	30	[-100,100]	0
$F_7(X) = \sum_{j=0}^{D} j{x_j}^4 + random[0,1]$	Quartic	30	[-128,128]	0

TABLE II. Description of Multimodal Benchmark Functions

Function	Description	Dim	Range	f_{min}
$F_8(X) = \sum_{j=1}^{D} -x_j \sin(\sqrt{\lvert x_j \rvert})$	Schwefel	30	[-500,500]	−418.98 29 ∗n
$F_9(X) = \sum_{j=1}^{D} [{x_j}^2 - 10\cos(2\pi x_j) + 10]$	Rastrigin	30	[−5.12,5.12]	0
$F_{10}(X) = -20\exp\left(-0.2\sqrt{\frac{1}{D}\sum_{j=1}^{D} {x_j}^2}\right) - \exp\left(\frac{1}{D}\sum_{j=1}^{D}\cos(2\pi x_j)\right) + 20 + e$	Ackley	30	[-32,32]	0

Function	Description	Dim	Range	f_{min}
$F_{11}(X) = \frac{1}{4000}\sum_{j=1}^{D} x_j{}^2 - \prod_{j=1}^{D}\cos\left(\frac{x_j}{\sqrt{j}}\right) + 1$	Griewank	30	[-600,600]	0
$F_{12}(X) = \frac{\pi}{D}\left\{10\sin(\pi y_1) + \sum_{j=1}^{D-1}(y_j-1)^2\left[1 + 10sin^2(\pi y_{j+1})\right] + (y_D-1)^2\right\} + \sum_{j=1}^{D} u(x_j,10,100,4)$ $y_j = 1+\frac{x_j+1}{4}$ $u(x_j,a,k,m) = \begin{cases} k(x_j-a)^m & x_j > \\ 0 & -a < x_j \\ k(-x_j-a)^m & x_j < \end{cases}$	Penalized	30	[-50,50]	0
$F_{13}(X) = 0.1\left\{sin^2(3\pi x_1) + \sum_{j=1}^{D}(x_j-1)^2\left[1+sin^2(3\pi x_j+1)\right] + (x_D-1)^2[1+sin^2(2\pi x_D)]\right\} + \sum_{j=1}^{D} u(x_j,5,100,4)$	Penalize 2	30	[-50,50]	0

TABLE III. DESCRIPTION OF FIXED-DIMENSION MULTIMODAL BENCHMARK FUNCTIONS

Function	Description	Dim	Range	f_{min}
$F_{14}(X) = (\frac{1}{500} + \sum_{j=1}^{25}\frac{1}{j+\sum_{k=1}^{2}(x_k-a_{kj})^6})^{-1}$	Foxholes	2	[-65,65]	1
$F_{15}(X) = \sum_{j=1}^{11}[a_j - \frac{x_1(b_j^2+b_jx_2)}{b_j^2+b_jx_3+x_4}]^2$	Kowalik	4	[−5,5]	0.0003
$F_{16}(X) = 4x_1^2 - 2.1x_1^2 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4$	Six-hump Camel-Back	2	[-5,5]	-1.0316
$F_{17}(X) = (x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6)^2 + 10(1-\frac{1}{8\pi})cos\,x_1 + 10$	Branin	2	[-5,5]	0.398
$F_{18}(X) = [1+(x_1+x_2+1)^2(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)] * [30 + (2x_1-3x_2)^2(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)]$	Goldstein-Price	2	[-2,2]	3
$F_{19}(X) = -\sum_{j=1}^{4} c_j \exp(-\sum_{k=1}^{3} a_{jk}(x_k - p_{jk})^2)$	Hartman 3	3	[1,3]	-3.86
$F_{20}(X) = -\sum_{j=1}^{4} c_j \exp(-\sum_{k=1}^{6} a_{jk}(x_k - p_{jk})^2)$	Hartman 6	6	[0,1]	-3.32
$F_{21}(X) = -\sum_{j=1}^{5}[(X-a_j)(X-a_j)^T + c_j]^{-1}$	Shekel 5	4	[0,10]	-10.1532
$F_{22}(X) = -\sum_{j=1}^{7}[(X-a_j)(X-a_j)^T + c_j]^{-1}$	Shekel 7	4	[0,10]	-10.4028
$F_{23}(X) = -\sum_{j=1}^{10}[(X-a_j)(X-a_j)^T + c_j]^{-1}$	Shekel 10	4	[0,10]	-10.5363

1) Comparison with SMA and SMA variants:

a) Performance analysis: This section compares the efficacy of MSMA to that of SMA [14] and SMA variants. Specifically, S^{2}SMA [25], ESMA [26], LSMA [19], and AOSMA [24] are executed based on the parameters shown in Table IV. The mean, and standard deviation of MSMA and other algorithms are reported in Table V. The ranking was determined by using an average of 30 runs. MSMA achieved the optimal value, zero, or the best result in most functions relative to other algorithms. This is due to the application of rules that aid in selecting the optimal algorithm, which in turn enables the achievement of optimal results. However, the outcomes were not satisfactory due to the nature of the functions F5, F6, F7, F19, and F20.

TABLE IV. CONFIGURATION PARAMETERS FOR THE EXAMINED ALGORITHMS

Method	Population size	The maximum number of iterations	Other parameters
MSMA (Proposed)	30	10^3	z = 0.03, μ =0.5, ϵ =0.02, α = 5 and β =3/2
SMA [14]	30	10^3	z = 0.03
LSMA [19]	30	10^3	z = 0.03

AOSMA [24]	30	10^3	$z = 0.03$
ESMA [26]	30	10^3	$z = 0.03$

TABLE V. RESULTS OF CEC 2005 FUNCTIONS

Function	Fitness	Algorithm					
		MSMA	***S²SMA***	***SMA***	***ESMA***	***LSMA***	***AOSMA***
F1	Mean	0.000E+00	6.525E-06	0.000E+00	0.000E+00	0.000E+00	0.000E+00
	Std	0.000E+00	2.863E-06	0.000E+00	0.000E+00	0.000E+00	0.000E+00
F2	Mean	0.000E+00	2.250E-03	1.233E-196	2.988E-227	0.000E+00	0.000E+00
	Std	0.000E+00	8.600E-04	0.000E+00	0.000E+00	0.000E+00	0.000E+00
F3	Mean	0.000E+00	7.737E-04	0.000E+00	0.000E+00	0.000E+00	0.000E+00
	Std	0.000E+00	3.599E-04	0.000E+00	0.000E+00	0.000E+00	0.000E+00
F4	Mean	0.000E+00	9.313E-02	6.527E-201	2.532E-295	0.000E+00	0.000E+00
	Std	0.000E+00	7.777E-02	0.000E+00	0.000E+00	0.000E+00	0.000E+00
F5	Mean	3.129E-02	8.328E-01	1.174E+00	9.674E-01	**1.866E-02**	2.254E-02
	Std	1.165E-01	4.461E+00	4.931E+00	4.676E+00	1.451E-02	6.369E-02
F6	Mean	4.397E-06	3.152E-05	9.883E-04	5.087E-04	1.710E-04	**3.987E-06**
	Std	2.498E-06	1.238E-05	4.196E-04	2.147E-04	6.327E-05	1.589E-06
F7	Mean	4.950E-05	3.833E-03	7.968E-05	7.292E-05	5.945E-05	**2.761E-05**
	Std	6.893E-05	4.007E-03	7.901E-05	7.692E-05	6.367E-05	2.676E-05
F8	Mean	**-1.257E+04**	-1.257E+04	-1.257E+04	-1.257E+04	-1.257E+04	-1.257E+04
	Std	1.563E-04	1.724E-04	8.373E-02	2.321E-02	4.497E-03	1.784E-04
F9	Mean	0.000E+00	3.574E-06	0.000E+00	0.000E+00	0.000E+00	0.000E+00
	Std	0.000E+00	1.111E-06	0.000E+00	0.000E+00	0.000E+00	0.000E+00
F10	Mean	8.882E-16	6.185E-04	8.882E-16	8.882E-16	8.882E-16	8.882E-16
	Std	0.000E+00	1.318E-04	0.000E+00	0.000E+00	0.000E+00	0.000E+00
F11	Mean	0.000E+00	5.694E-03	0.000E+00	0.000E+00	0.000E+00	0.000E+00
	Std	0.000E+00	1.168E-02	0.000E+00	0.000E+00	0.000E+00	0.000E+00
F12	Mean	1.193E-06	**7.800E-07**	8.589E-04	4.542E-04	3.119E-05	3.051E-05
	Std	7.749E-07	3.096E-07	1.335E-03	5.612E-04	1.612E-05	1.576E-04
F13	Mean	**9.972E-06**	1.106E-05	7.211E-04	4.259E-04	2.254E-04	1.110E-03
	Std	5.603E-06	4.735E-06	4.926E-04	2.342E-04	1.175E-04	3.352E-03
F14	Mean	**9.980E-01**	9.980E-01	9.980E-01	9.980E-01	9.980E-01	9.980E-01
F15	Mean	**3.448E-04**	4.613E-04	5.015E-04	4.679E-04	5.629E-04	5.246E-04
	Std	8.868E-05	3.506E-04	2.224E-04	1.964E-04	2.814E-04	3.570E-04
F16	Mean	**-1.032E+00**	-1.032E+00	-1.032E+00	-1.032E+00	-1.032E+00	-1.032E+00
	Std	6.421E-13	2.593E-11	1.878E-10	3.863E-10	1.630E-09	5.656E-12
F17	Mean	**3.979E-01**	3.979E-01	3.979E-01	3.979E-01	3.979E-01	3.979E-01
	Std	2.171E-09	1.531E-07	3.122E-08	4.086E-08	7.897E-08	5.215E-09
F18	Mean	3.000E+00	3.000E+00	3.000E+00	3.000E+00	3.000E+00	3.000E+00
	Std	3.720E-10	1.187E-09	6.592E-12	9.030E-13	7.936E-08	7.377E-10
F19	Mean	-3.863E+00	-3.863E+00	**-3.863E+00**	-3.863E+00	-3.863E+00	-3.863E+00
	Std	3.073E-06	1.824E-05	3.073E-08	8.330E-07	2.485E-06	9.483E-07
F20	Mean	-3.254E+00	-3.230E+00	-3.239E+00	-3.231E+00	-3.237E+00	**-3.270E+00**
	Std	6.014E-02	5.138E-02	5.543E-02	5.133E-02	5.636E-02	6.038E-02
F21	Mean	**-1.015E+01**	-1.015E+01	-1.015E+01	-1.015E+01	-1.015E+01	-1.015E+01
	Std	5.139E-07	9.000E-05	9.021E-05	1.180E-04	3.903E-05	5.340E-06
F22	Mean	**-1.040E+01**	-1.040E+01	-1.040E+01	-1.040E+01	-1.040E+01	-1.040E+01
	Std	7.056E-07	7.015E-05	9.711E-05	1.015E-04	3.713E-05	6.514E-06
F23	Mean	**-1.054E+01**	-1.054E+01	-1.054E+01	-1.054E+01	-1.054E+01	-1.054E+01
	Std	1.544E-07	7.791E-05	1.054E-04	1.122E-04	4.224E-05	6.612E-06

b) Analysis of execution time: The computer's software and hardware specifications used to conduct the investigations in this study are elaborated on in Table VI. Table VII displays the computational time (in seconds) for three different algorithms: MSMA, SMA, and AOSMA. According to Table VIII, the SMA algorithm achieved a computation time of 0.51318182 seconds, while the AOSMA algorithm recorded a shorter time at 0.260618 seconds. As a result of hybridizing AOSMA and S^2SMA, MSMA achieved a computation time of 0.380505 seconds and thus outperformed the original SMA algorithm. These results indicate that MSMA shows promise in improving task-specific computational time compared to the traditional SMA approach. Notably, the programming language, programmer proficiency, and machine configuration influence the CPU time utilized by each method.

TABLE VI. SETTING INFORMATION FOR HARDWARE AND SOFTWARE

Item	Component	Setting
Hardware	CPU	Intel(R) Core (TM) i7-10700
	Frequency	2.9 GHz
	RAM	16GB
	GPU	Nvidia GeForce GTX 1660 Super
	SSD	256 GB
	Hard Drive	2 TB
Software	Operating system	Windows 10
	Language	MATLAB R2021a

TABLE VII. COMPUTATIONAL TIME ANALYSIS

	MSMA (Proposed)	SMA	AOSMA
Time (Second)	0.380505	0.51318182	0.260618

2) Comparison with conventional algorithms:

a) Performance analysis: This section compares the performance of the MSMA algorithm with six popular metaheuristic algorithms: WOA [5], Multi-Verse Optimizer (MVO) [47], Grey Wolf Optimizer (GWO) [48], Sine Cosine Algorithm (SCA) [49], Arithmetic Optimization Algorithm (AOA) [50], and PSO [4] across unimodal and multimodal functions (F1-F13). The primary parameter configurations of these algorithms are displayed in Table VIII below. It has been demonstrated that the MSMA variant outperforms the original SMA and other SMA variants. Therefore, the upcoming comparative experiment will not include SMA for comparison.

According to Table IX, MSMA's ability to achieve highly competitive best fitness values frequently converges to zero or near-zero fitness on unimodal functions such as F1 and F2, emphasizing its exceptional exploitation efficiency. Moreover, on multimodal functions like F13, MSMA exhibits worthy exploration capabilities, navigating intricate landscapes and converging to optimal solutions. These findings collectively highlight MSMA as an effective metaheuristic algorithm with the potential for solving real-world problems in various domains.

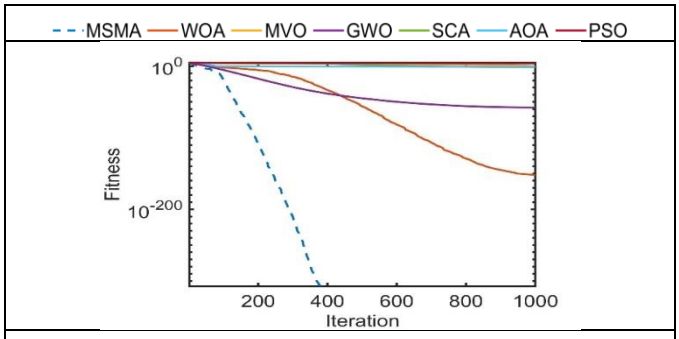
TABLE VIII. THE SETTING OF ALGORITHMS' PARAMETERS

Method	Population size	The maximum number of iterations	Other parameters
MSMA (Proposed)	30	10^3	z = 0.03, μ =0.5, ϵ =0.02, α = 5 and β =3/2
(WOA) [5]	30	10^3	$a1 = 2-0$; $a2 = -1--2$; $b = 1$
(MVO) [47]	30	10^3	Wormhole Existence Probability WEPMax = 1; WEPMin = 0.2;
(GWO) [48]	30	10^3	a: 2−1
(SCA) [49]	30	10^3	$a = 2$
AOA [50]	30	10^3	μ =0.5 and α = 5
(PSO) [4]	30	10^3	c1 = 2.5 – 0.5, c2 = 0.5–2.5, w = 0.9–0.5.

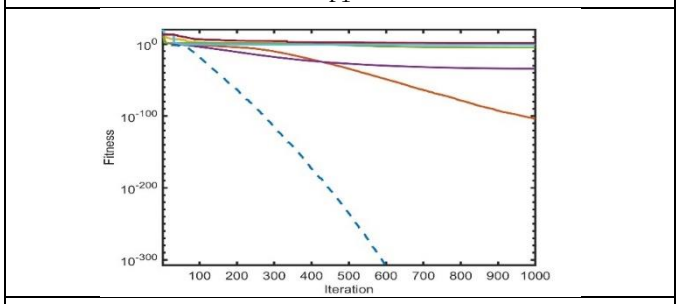
TABLE IX. COMPARISON MSMA WITH CONVENTIONAL ALGORITHMS

Function	Fitness	Algorithm						
		MSMA	***WOA***	***MO***	***GWO***	***SCA***	***AOA***	***PSO***
F1	Mean	**0**	6.9E-153	0.313323	1.81E-58	0.078575	2.23E-32	5866.422
	Std	0	3.3E-152	0.09829	7.72E-58	0.23702	1.22E-31	1311.936
F2	Mean	**0**	4E-104	0.42004	8.46E-35	3.22E-05	**0**	48.39758
	Std	0	1.9E-103	0.093926	1.14E-34	6.67E-05	0	10.43552
F3	Mean	**0**	23143.08	44.96597	1.64E-15	3319.778	0.004737	24646.9
	Std	0	9726.587	20.69163	6.89E-15	2581.694	0.009056	5292.921
F4	Mean	**0**	41.81578	0.899134	2.87E-14	21.89053	0.024824	37.87362
	Std	0	31.17343	0.283396	1.05E-13	12.17732	0.020985	3.295181
F5	Mean	**0.85445**	27.21463	204.3283	26.77613	184.4234	28.26249	2735175
	Std	4.564721	0.553541	224.4148	0.829596	300.5298	0.40652	1167378
F6	Mean	**5.59E-06**	0.093981	0.315247	0.555492	4.81354	2.778531	6032.94
	Std	5.29E-06	0.116612	0.079513	0.329167	0.783288	0.27915	1379.102
F7	Mean	9.38E-05	0.002088	0.021331	0.000915	0.035331	**3.45E-05**	2.15723
	Std	8.94E-05	0.00177	0.008588	0.000562	0.030723	3.86E-05	0.798218
F8	Mean	**-12569.5**	-11210.8	-7972.48	-6055.25	-3921.06	-5738.88	-3277.07
	Std	0.00015	1489.677	621.0526	956.9094	261.2019	492.3883	419.031
F9	Mean	**0**	1.89E-15	109.1013	0.848748	25.83291	**0**	267.1143
	Std	0	1.04E-14	30.77091	2.414291	37.67391	0	19.37847
F10	Mean	**8.88E-16**	4.56E-15	1.278301	1.6E-14	13.02186	**8.88E-16**	13.72874
	Std	0	2.38E-15	0.96858	2.79E-15	8.510519	0	0.871003
F11	Mean	**0**	**0**	0.568299	0.001077	0.177192	0.090559	57.97914
	Std	0	0	0.091516	0.003314	0.204652	0.067312	11.14415
F12	Mean	**1.19E-06**	0.005829	1.557851	0.040531	2.313704	0.414541	653080.9
	Std	6.37E-07	0.0047	1.298217	0.019803	3.730789	0.04977	636192.7
F13	Mean	**3.09E-05**	0.192608	0.073073	0.521447	12.55231	2.78815	5119619
	Std	0.00011	0.115827	0.037876	0.208239	35.92383	0.095486	3507365

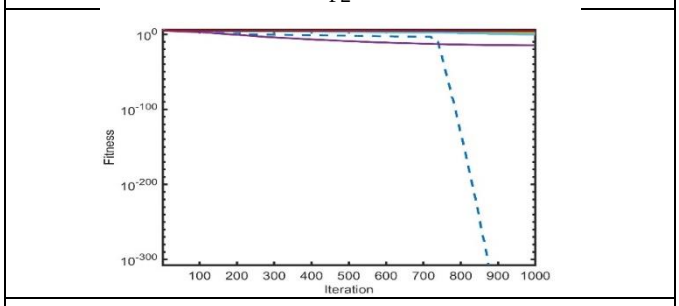
b) Convergence curve: In this section, Fig. 6 shows the convergence curves of MSMA compared to WOA [5], MVO [47], GWO [48], SCA [49], AOA [50], and PSO [4]. Fig. 6 displays convergence curves derived from the average best objective function value achieved over 30 runs, as detailed in Table IX. The x-axis represents 1000 iterations, while the y-axis represents the maximum score achieved. The results show that MSAM is superior to its competitors in most of the unimodal functions (1–7), and this reflects its high ability in the exploitation phase. Furthermore, MSAM's exploratory capabilities were showcased in multimodal functions (8–13), highlighting its superiority in all functions. Overall, it demonstrates that the convergence of MSMA is significantly superior to that of other algorithms across most functions. This is due to VSRR, which intelligently switches between algorithms in MSMA to take advantage of its exploitation and exploration capabilities.



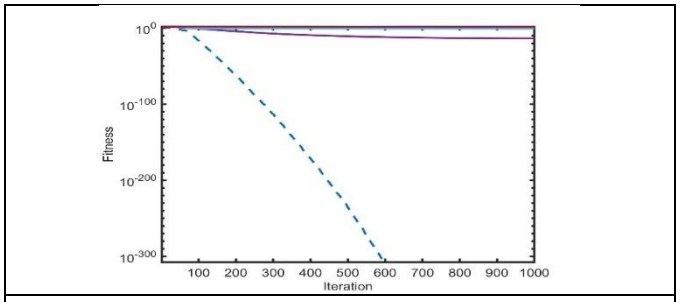
F1



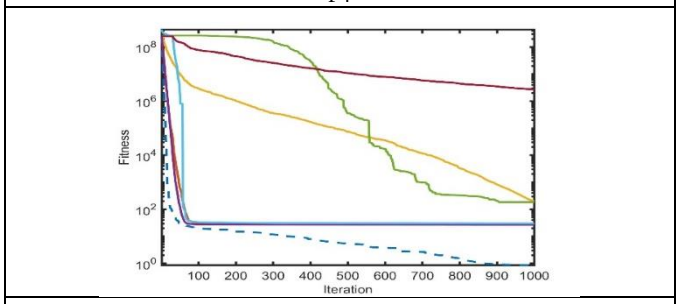
F2



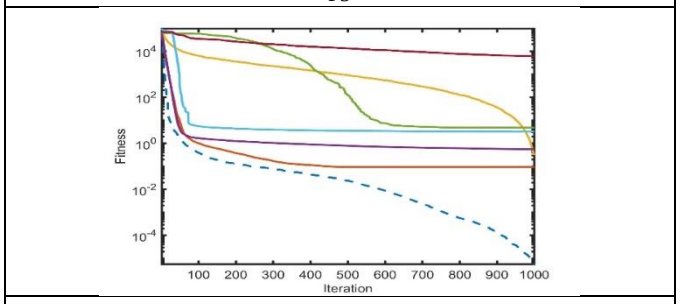
F3



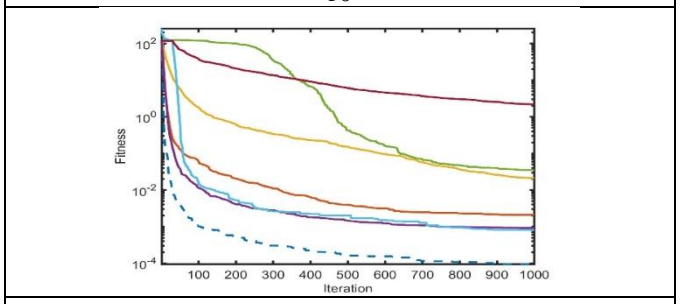
F4



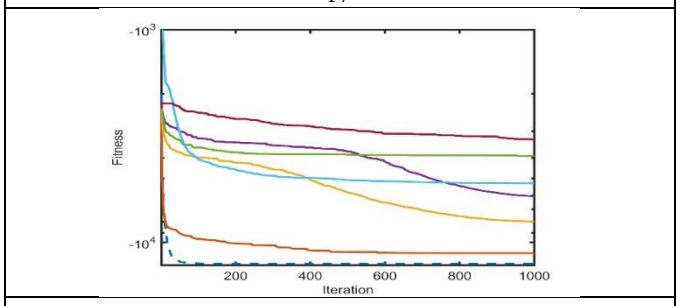
F5



F6



F7



F8

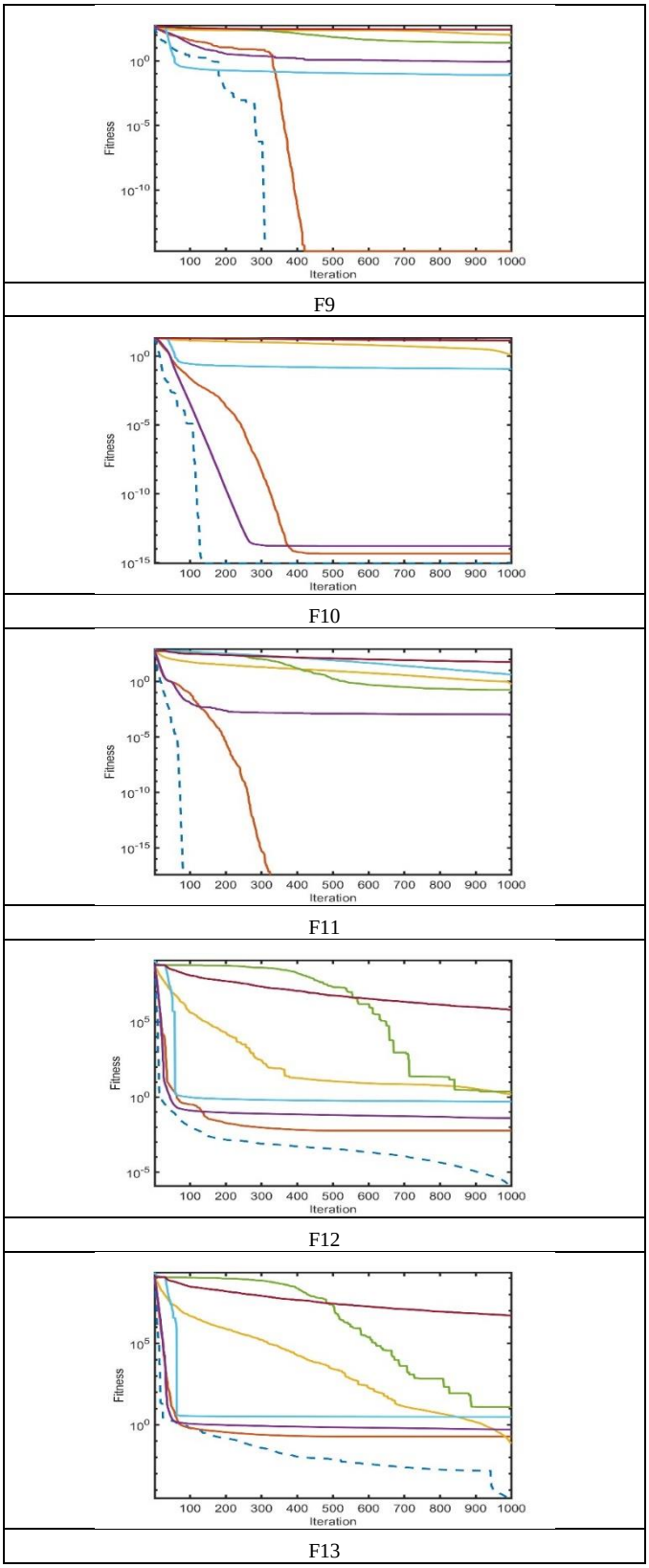


Fig. 6. The convergence curves for unimodal and multimodal functions.

B. Experimental Results on Constraints Problems (Engineering Design Problems)

The performance of MSMA was assessed by applying the method to solve various engineering design problems. These included a cantilever beam problem, a welded beam design problem, a pressure vessel problem, a compression coil spring design problem, a multiple disc clutch brake problem, a speed reducer problem, and a gear train design problem. The mathematical formulas relating to these problems are provided in "Appendix A". These validations evaluated the effectiveness and suitability of MSMA in tackling different design challenges.

This experiment standardized the parameters for all optimization techniques to ensure a fair comparison. The maximum number of function iterations was set to 10,000, and the population size was set to 30. For statistically reliable results, each method underwent 30 runs independently.

1) Performance analysis: This section compares the efficacy of MSMA to that of SMA [14] and SMA variants (S^2SMA [25], LSMA [36], AOSMA [24], and ESMA [26]). Table X shows performance metrics for MSMA and the other algorithms on Engineering design problems, including the mean, and the standard deviation.

Based on the obtained results, in the problem of Cantilever Beam analysis, MSMA, boasting a mean of 13.36523309, clearly outperforms its counterparts. ESMA, LSMA, and SMA yield closely clustered means of 13.36532, 13.36531551, and 13.36536, respectively, while AOSMA displays a slightly elevated average. This highlights MSMA's superior effectiveness. Likewise, in Welded Beam problem assessments, MSMA's mean of 1.724885178 is notably superior to its peers. ESMA closely trails with a mean of 1.724979, while other algorithms register marginally higher averages, underscoring the unmistakable dominance of MSMA in this context. Transitioning to the Pressure Vessel problem, MSMA stands out as the top-performing algorithm. Its mean of 6766.643344 significantly outperforms SMA, ESMA, and LSMA, all of which yield notably higher means. This underscores MSMA's exceptional suitability for this specific problem. In the Compression Coil Spring Design problem, MSMA's mean of 0.01284604 distinctly outshines alternative algorithms, which yield significantly higher averages. This glaring disparity underscores the exceptional performance of MSMA in this scenario.

Moreover, in the Multiple Disk Clutch Brake problem, MSMA, SMA, and ESMA exhibit closely aligned means, with MSMA marginally leading. While LSMA and AOSMA register slightly higher values, MSMA's marginal lead implies superior efficacy for this problem. In the Speed Reducer problem, MSMA, ESMA, and AOSMA stand out with proximate mean values, with MSMA in the lead. In contrast, SMA and LSMA yield substantially higher averages, reinforcing the notable performance of MSMA. Lastly, in the Gear Train Design problem, MSMA's mean value of 3.25763E-20 is strikingly lower than those of alternative algorithms, which produce considerably higher means, unequivocally solidifying its unparalleled suitability for this specific function. These results demonstrate MSMA's superior performance across various engineering problems, affirming its pivotal role in optimization endeavours.

TABLE X. STATISTICAL RESULTS OF ENGINEERING DESIGN PROBLEMS

Function	Fitness	Algorithm					
		MSMA	***S²SMA***	***SMA***	***ESMA***	***LSMA***	***AOSMA***
Cantilever beam	Mean	**13.36523**	13.36534	13.36536	13.36532	13.36532	13.36526
	Std	2.23E-05	0.0001	7.36E-05	9.2E-05	5.69E-05	3.74E-05
Cantilever beam	Mean	**1.724885**	1.725027	1.725076	1.724979	1.725094	1.725079
	Std	7.13E-05	0.000174	0.000276	0.000113	0.00021	0.000229
Pressure Vessel	Mean	**6766.643**	6861.241	7818 1.51	6880 8.52	7491 9.22	3755 4.12
	Std	565.1678	501.2	4022 6.18	4486 6.76	4230 2.85	4491 2.3
Compression Coil Spring design	Mean	**0.012846**	0.013245	5000 0.01	4000 0.01	5333 3.34	5666 6.67
	Std	0.000151	0.000301	5085 4.76	4982 7.28	5074 1.62	5040 0.69
Multiple disk clutch brake	Mean	**0.25977**	0.259784	0.259774	0.259774	0.259785	0.259771
	Std	3.18E-06	1.26E-05	2.98E-06	4.5E-06	1.47E-05	1.34E-06
Speed reducer	Mean	**2996.351**	2996.352	1000 00	9676 6.54	9353 3.09	9676 6.54
	Std	0.009292	0.003918	0	1771 0.36	2461 0.61	1771 0.36
Gear train design	Mean	**3.26E-20**	2E-14	3.74E-14	4.35E-14	3.13E-14	5.45E-15
	Std	8E-20	2.84E-14	1.31E-13	1.08E-13	6.49E-14	1.44E-14

2) Statistical analysis: To statistically evaluate the performance of MSMA and the compared algorithms, including SMA [14] S and SMA variants (S^2SMA [25], LSMA [36], AOSMA [24], and ESMA [26]), on various engineering design problems. Calculating the p-value of the Wilcoxon signed-rank test [51]. Each value greater than 0.05 is displayed in bold font, indicating that the difference is not statistically significant. The calculated p-values indicate substantial evidence of differentiation, as shown in Table XI MSMA's p-values are smaller than 0.05 in the majority of cases, indicating significant differences. Notably, the "Pressure Vessel" problem exhibits a relatively large p-value (approximately 0.76) when comparing MSMA to AOSMA, indicating that the difference with AOSMA is not statistically significant. In contrast, for the "Gear train design" problem, the p-values consistently indicate significant differences, indicating that MSMA outperforms all compared algorithms. These results demonstrate the superior performance of MSMA and its potential as an efficient optimization method for complex engineering design problems.

TABLE XI. P-VALUES FOR MSMA VERSUS OTHER COMPETITORS ON ENGINEERING DESIGN PROBLEMS

Function	MSMA vs. S^2SMA	MSMA vs. SMA	MSMA vs. ESMA	MSMA vs. LSMA	MSMA vs. AOSMA
Cantilever beam	7.77255E-09	7.38029E-10	2.37682E-07	2.19474E-08	0.000471375
Welded Beam	3.64589E-08	1.42942E-08	1.15665E-07	1.10234E-08	2.83145E-08
Pressure Vessel	**0.157975689**	2.66709E-06	0.000244046	0.000377215	**0.761297126**
Compression Coil Spring design	1.8731E-07	0.005708009	**0.619007153**	0.031019514	0.001643841
Multiple disk clutch brake	8.10136E-10	3.96477E-08	7.69496E-08	7.38029E-10	7.59915E-07
Speed reducer	1.35943E-07	1.21178E-12	4.21155E-12	3.68819E-12	8.15959E-12
Gear train design	3.01986E-11	7.38029E-10	3.01986E-11	3.01986E-11	3.01986E-11

3) Comparison with conventional algorithms: This section aims to evaluate the performance of MSMA through a comprehensive comparison with six popular metaheuristic algorithms: WOA [5], MVO [47], GWO [48], SCA [49], AOA [50], and PSO [4]. The comparison is conducted across seven distinct engineering design problems to thoroughly assess their capabilities in solving engineering problems. The main parameter settings for each algorithm are outlined in Table VIII.

Beginning with the Cantilever Beam Design Problem, the analysis reveals that MSMA exhibits competitive performance, achieving optimal values for variables (x1 to x5) and an optimal cost of 13.36520828, as demonstrated in Table XII. This outcome underscores the effectiveness of MSMA in addressing structural engineering challenges, where precise optimization is paramount for ensuring structural integrity and efficiency.

Similarly, in the Welded Beam Problem, MSMA demonstrates notable performance with an optimal cost of 1.724852759, as presented in Table XIII, indicating its capability to navigate the complexities inherent in welding design optimization. The results further validate the robustness of MSMA in handling diverse engineering scenarios, where intricate design considerations must be balanced to achieve optimal outcomes.

The Pressure Vessel Problem, as presented in Table XIV, further emphasizes the diversity of MSMA's capabilities. It showcases optimal values for variables and an optimal cost of 5885.332794. It highlights MSMA's adaptability to multifaceted challenges in pressure vessel design optimization, where complex geometrical and operational constraints influence the design space.

In the Compression Coil Spring Design Problem, MSMA continues to demonstrate competitive results, achieving an optimal cost of 0.012665319, as presented in Table XV. This performance highlights the efficacy of MSMA in optimizing mechanical components, where precision in design parameters is crucial for achieving desired spring characteristics and performance metrics.

Table XVI illustrates the optimal values for variables (x1, x2, x3, x4, x5) and their respective optimal costs achieved by various algorithms in the Multiple Disk Clutch Brake scenario. MSMA outperforms competitors by attaining an optimal cost of 0.259768995. In contrast, other algorithms exhibit slightly different values for the variables. It highlights the effectiveness of MSMA in this context.

Similarly, Table XVII provides a comparative analysis for the Speed Reducer Problem, where MSMA excels in achieving an optimal cost of 2996.348166. Competing algorithms, on the other hand, are unable to reach the same degree of accuracy. MSMA's reliability and effectiveness are demonstrated by its ability to handle the complexity of this problem.

In the context of gear train design optimization, Table XVIII highlights the effectiveness of the MSAM algorithm with an optimal cost of 4.29529E-26. Additionally, WOA achieves a noteworthy optimal cost of 0, emphasizing its competitive performance. These findings underscore the capabilities of MSAM and WOA in addressing complex engineering optimization challenges.

TABLE XII. COMPARISON RESULTS OF THE CANTILEVER BEAM DESIGN PROBLEM

Algorithms	Optimal values for variables					Optimal cost
	x1	*x2*	*x3*	*x4*	*x5*	
MSMA	6.017085383	5.311288687	4.488476538	3.507273784	2.149604785	**13.36520828**
WOA	5.700449107	5.397613593	4.814600886	3.522731364	2.119436619	13.38920896
MVO	6.033559535	5.307530 25	4.440377082	3.528653236	2.165715561	13.36529853
GWO	6.030295996	5.311933104	4.485345216	3.494561661	2.151681834	13.36520866
SCA	6.256390441	6.108302 51	4.446189485	3.086518141	2.009635551	13.44560967
AOA	6.314864496	5.699730447	3.986353036	3.898579909	2.105572722	13.49454764
PSO	5.716182614	5.394840612	4.923773757	3.426754426	2.132218051	13.39554248

TABLE XIII. COMPARISON RESULTS OF THE WELDED BEAM PROBLEM

Algorithms	Optimal values for variables				Optimal cost
	x1	*x2*	*x3*	*x4*	
MSMA	0.20572668	3.470554907	9.036623951	0.205729641	**1.724852759**
WOA	0.416992489	2.031718826	6.338092508	0.421197407	1.757628876
MVO	0.204572029	3.495334409	9.042511959	0.205708116	1.725770029
GWO	0.205483433	3.475777453	9.036731042	0.205741147	1.724921656
SCA	0.206779645	3.346727647	9.463618339	0.210822386	1.744207622
AOA	0.208707168	3.156291235	10	0.247529632	1.854111628
PSO	0.211161682	3.416026615	8.890902048	0.21255867	1.733532831

TABLE XIV. COMPARISON RESULTS OF THE PRESSURE VESSEL PROBLEM

Algorithms	Optimal values for variables				Optimal cost
	x1	*x2*	*x3*	*x4*	
MSMA	1.258828444	0.622239552	65.22427172	10.00413828	**5885.332794**
WOA	74.39811474	34.47896969	46.42166663	73.12900591	5913.484457
MVO	89.65475933	74.80944321	18.909546	166.6995299	6432.102507
GWO	0.77873673	0.385013082	40.34876174	199.595325	5886.112827
SCA	0.799649958	0.428055843	40.71461012	200	5968.711993
AOA	34.09159014	87.13884443	13.35195248	51.93686087	9424.698317
PSO	36.82855393	79.53717197	52.42213373	69.31768329	6155.484164

TABLE XV. COMPARISON RESULTS OF THE COMPRESSION COIL SPRING DESIGN PROBLEM

Algorithms	Optimal values for variables			Optimal cost
	x1	*x2*	*x3*	
MSMA	0.055584231	0.457867024	7.138747128	**0.012665319**
WOA	0.059352736	0.570654818	4.793787974	0.012672374
MVO	0.057411627	0.510629027	5.858130197	0.012702184
GWO	0.030415911	0.746376689	2.882659855	0.01266583
SCA	0.049565332	0.307684113	15	0.012751116
AOA	0.076649751	1.3	2	0.015289034
PSO	0.050062779	0.315733397	14.66722903	0.012701516

TABLE XVI. MULTIPLE DISK CLUTCH BRAKE

Algorithms	Optimal values for variables					Optimal cost
	x1	*x2*	*x3*	*x4*	*x5*	
MSMA	69.9999999	90.00000002	1.000000004	1000	2.312782041	**0.259768995**
WOA	70	90	1	1000	2.312782578	0.259769039
MVO	70.00153892	90.00180212	1	999.7168915	2.313482158	0.259781877
GWO	69.99852945	90	1	1000	2.312864645	0.259774817
SCA	69.63372395	90	1	1000	2.347561762	0.260725078
AOA	80	100.7047151	1	1000	2.32781793	0.277270665
PSO	69.99879821	90	1	1000	2.312844506	0.259783161

TABLE XVII. COMPARISON RESULTS OF THE SPEED REDUCER PROBLEM

Algorithms	Optimal values for variables							Optimal cost
	x1	*x2*	*x3*	*x4*	*x5*	*x6*	*x7*	
MSMA	3.500000002	0.7	17	7.300000013	7.80000075	3.350214675	5.286683234	**2996.348166**
WOA	3.5	0.7	17	8.086052026	8.061415721	3.356439258	5.348209953	3001.995774

MVO	3.501974802	0.7	17	7.405570651	8.070442656	3.353599328	5.286782167	2998.474657
GWO	3.500234799	0.7	17.00022304	7.323444697	7.801214073	3.350567429	5.286801157	2996.877194
SCA	3.551329239	0.7	17	7.781838931	8.3	3.422479892	5.310732943	3034.002185
AOA	3.6	0.7	17	7.3	8.3	3.516263029	5.294372667	3074.222921
PSO	2.626401315	0.728946859	20.50362653	8.249249684	8.218561811	3.380030355	5.348181951	2997.389625

TABLE XVIII. COMPARISON RESULTS OF THE GEAR TRAIN DESIGN PROBLEM

Algorithms	**Optimal values for variables**				**Optimal cost**
	x1	***x2***	***x3***	***x4***	
MSMA	20.51707896	14.28150743	12	57.89428103	4.29529E-26
WOA	56.38376759	12.21452818	33.2686516	49.95209432	**0**
MVO	18.0025215	12.49738546	12	57.73816517	1.03484E-18
GWO	50.71522562	17.09261159	24.15294502	56.42037706	2.93402E-17
SCA	58.2894748	40.59354904	12	57.92398093	5.63401E-16
AOA	59.97591416	12.00002286	43.26335633	60	2.99093E-15
PSO	42.42788592	30.55243375	12	59.89162803	2.5461E-15

C. Discussion

The Merged Slime Mould Algorithm (MSMA) results demonstrate its effectiveness across benchmark functions and engineering design problems. Evaluating 23 continuous benchmark functions from the CEC 2005 revealed that MSMA excels in achieving optimal results, particularly in unimodal functions where exploitation is crucial. Its performance in multimodal functions illustrates robust exploration capabilities, effectively navigating complex landscapes and avoiding local optima.

Comparisons with other Slime Mould Algorithm (SMA) variants and established metaheuristic algorithms like WOA, GWO, and PSO showed that MSMA consistently outperforms its peers. The mean and standard deviation metrics analysis highlight MSMA's ability to frequently achieve optimal or near-optimal fitness values. The convergence curves indicate that MSMA delivers rapid convergence, leveraging Vertical Smart Switching Rules (VSRR) for intelligent algorithm switching, thus enhancing both exploitation and exploration strategies.

MSMA's superiority in engineering design problems is further validated. For instance, the Cantilever Beam problem achieved significantly lower mean values compared to other algorithms. Similar trends were noted in the Welded Beam and Pressure Vessel problems, with statistical significance confirmed through the Wilcoxon signed-rank test. These results underscore MSMA's reliability and efficiency in tackling complex engineering challenges.

The promising outcomes of MSMA open several exciting avenues for future research. Exploring hybridization techniques that combine MSMA with advanced optimization algorithms could further enhance its performance. Additionally, adapting Vertical Smart Switching Rules (VSRR) for dynamic problem landscapes may improve efficiency. Future studies could also validate MSMA through real-world case studies, ensuring its practical applicability across diverse industries. Such explorations would significantly contribute to the optimization field and enhance MSMA's utility in addressing complex challenges, instilling a sense of optimism and hope for its continuous improvement.

V. CONCLUSION

In conclusion, this paper introduced MSMA as a dynamic hybridization approach engineered to significantly enhance the performance of the traditional SMA in tackling low-dimensional optimization problems compared to other algorithms. The proposed technique merges two existing SMA variants, AOSMA and S^2SMA, through the incorporation of embedded Vertical Smart Switching Rules (VSSR). VSSR enables dynamic switching between algorithms based on problem-specific attributes, thereby boosting adaptability and operational efficiency. The MSMA's unique integration strategy eliminates the need for multiple algorithm initializations as well as avoids the need for memory-based switching. Instead, it relies on adaptive and intelligent switching rules to exploit the strengths of both algorithms. This represents a notable advancement compared to previous integrations of SMA.

The proposed MSMA has been fully validated on ten real-world engineering challenges and basic benchmark problems CEC 2005 using statistical and numerical analyses. The experimental results highlight MSMA's superiority over current approaches and demonstrate its potential to provide innovative solutions for complex engineering designs. This study provides additional evidence that MSMA consistently achieves the highest mean fitness values and shows the fastest rates of convergence among the algorithms evaluated, demonstrating its superior performance in addressing engineering design problems. Remarkably, when compared to SMA, MSMA also showed improved computational efficiency, particularly in the Cantilever Beam problem. The Wilcoxon signed-rank test has statistically validated MSMA's outstanding performance in a variety of engineering problems, confirming its superiority and efficacy in resolving complex engineering design problems.

These findings validate the proposed MSMA's superiority over existing techniques, showcasing its potential to provide promising solutions for complex engineering design problems. Future research directions could pivot towards enhancing the VSSR mechanism to further improve MSMA's adaptability and robustness. Moreover, extending the exploration to other problem domains and conducting comparative studies with other state-of-the-art optimization algorithms would yield additional insights, paving the way for further advancements in optimization technology.

ACKNOWLEDGMENT

Conceptualization, S.A.S. and H.S.; Methodology, K.M.A.; Software, K.M.A.; Validation, S.A.S. and H.S.; Formal analysis, K.M.A.; Investigation, K.M.A.; Resources, K.M.A.; Data curation, K.M.A.; Writing – original draft, K.M.A.; Writing – review & editing, K.M.A. and H.S.; Visualization, S.A.S. and

H.S.; Supervision, S.A.S. and H.S.; Project administration, S.A.S. and H.S.; Funding acquisition, S.A.S. All authors have read and agreed to the published version of the manuscript.

REFERENCES

[1] M. H. Nadimi-Shahraki, S. Taghian, S. Mirjalili, H. Zamani, and A. Bahreininejad, "GGWO: Gaze cues learning-based grey wolf optimizer and its applications for solving engineering problems," J. Comput. Sci., vol. 61, no. June 2021, p. 101636, 2022.

[2] Y. Duan and X. Yu, "A collaboration-based hybrid GWO-SCA optimizer for engineering optimization problems," Expert Syst. Appl., vol. 213, no. PB, p. 119017, 2023.

[3] S. Chauhan and G. Vashishtha, "A synergy of an evolutionary algorithm with slime mould algorithm through series and parallel construction for improving global optimization and conventional design problem," Eng. Appl. Artif. Intell., vol. 118, no. December 2022, p. 105650, Feb. 2023.

[4] J. Kennedy and R. Eberhart, "Particle swarm optimization," in Proceedings of ICNN'95-international conference on neural networks, 1995, vol. 4, pp. 1942–1948.

[5] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," Adv. Eng. Softw., vol. 95, pp. 51–67, 2016.

[6] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," Futur. Gener. Comput. Syst., vol. 97, pp. 849–872, 2019.

[7] W. Michiels, E. H. L. Aarts, and J. Korst, Theoretical aspects of local search, vol. 13. Springer, 2007.

[8] F. Glover, "Tabu search—part I," ORSA J. Comput., vol. 1, no. 3, pp. 190–206, 1989.

[9] S. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi, "Optimization by simulated annealing," Science (80-.)., vol. 220, no. 4598, pp. 671–680, 1983.

[10] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," IEEE Trans. Evol. Comput., vol. 1, no. 1, pp. 67–82, 1997.

[11] X. Zhong, Z. You, and P. Cheng, "A hybrid optimization algorithm and its application in flight trajectory prediction," Expert Syst. Appl., vol. 213, no. PB, p. 119082, 2023.

[12] Q. S. Hamad, H. Samma, S. A. Suandi, and J. Mohamad-Saleh, "Q-learning embedded sine cosine algorithm (QLESCA)," Expert Syst. Appl., vol. 193, no. November 2021, p. 116417, 2022.

[13] H. Samma, J. Mohamad-Saleh, S. A. Suandi, and B. Lahasan, "Q-learning-based simulated annealing algorithm for constrained engineering design problems," Neural Comput. Appl., vol. 32, no. 9, pp. 5147–5161, 2020.

[14] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, "Slime mould algorithm: A new method for stochastic optimization," Futur. Gener. Comput. Syst., vol. 111, pp. 300–323, Oct. 2020.

[15] J. M. Abdullah and T. Ahmed, "Fitness Dependent Optimizer: Inspired by the Bee Swarming Reproductive Process," IEEE Access, vol. 7, pp. 43473–43486, 2019.

[16] V. Hayyolalam and A. A. P. Kazem, "Black widow optimization algorithm: a novel meta-heuristic approach for solving engineering optimization problems," Eng. Appl. Artif. Intell., vol. 87, p. 103249, 2020.

[17] L. Abualigah, M. Abd Elaziz, P. Sumari, Z. W. Geem, and A. H. Gandomi, "Reptile Search Algorithm (RSA): A nature-inspired meta-heuristic optimizer," Expert Syst. Appl., vol. 191, p. 116158, 2022.

[18] T. Kundu and H. Garg, "LSMA-TLBO: A hybrid SMA-TLBO algorithm with lévy flight based mutation for numerical optimization and engineering design problems," Adv. Eng. Softw., vol. 172, no. May, p. 103185, 2022.

[19] M. K. Naik, R. Panda, and A. Abraham, "Normalized square difference based multilevel thresholding technique for multispectral images using leader slime mould algorithm," J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 7, pp. 4524–4536, Jul. 2022.

[20] H. Lin et al., "Adaptive slime mould algorithm for optimal design of photovoltaic models," Energy Sci. Eng., vol. 10, no. 7, pp. 2035–2064, 2022.

[21] M. H. Hassan, S. Kamel, L. Abualigah, and A. Eid, "Development and application of slime mould algorithm for optimal economic emission dispatch," Expert Syst. Appl., vol. 182, no. May, p. 115205, 2021.

[22] S. Zhao et al., "Multilevel threshold image segmentation with diffusion association slime mould algorithm and Renyi's entropy for chronic obstructive pulmonary disease," Comput. Biol. Med., vol. 134, no. April, p. 104427, 2021.

[23] A. A. Ewees et al., "Improved Slime Mould Algorithm based on Firefly Algorithm for feature selection: A case study on QSAR model," Eng. Comput., no. 0123456789, 2021.

[24] M. K. Naik, R. Panda, and A. Abraham, "Adaptive opposition slime mould algorithm," Soft Comput., vol. 25, no. 22, pp. 14297–14313, 2021.

[25] K. M. Alhashash, H. Samma, and S. A. Suandi, "Fine-Tuning of Pre-Trained Deep Face Sketch Models Using Smart Switching Slime Mold Algorithm," Appl. Sci., vol. 13, no. 8, p. 5102, Apr. 2023.

[26] M. K. Naik, R. Panda, and A. Abraham, "An entropy minimization based multilevel colour thresholding technique for analysis of breast thermograms using equilibrium slime mould algorithm," Appl. Soft Comput., vol. 113, p. 107955, 2021.

[27] H. Chen, X. Li, S. Li, Y. Zhao, and J. Dong, "Improved Slime Mould Algorithm Hybridizing Chaotic Maps and Differential Evolution Strategy for Global Optimization," IEEE Access, vol. 10, no. May, pp. 66811–66830, 2022.

[28] A. A. Bhandakkar and L. Mathew, "Merging slime mould with whale optimization algorithm for optimal allocation of hybrid power flow controller in power system," J. Exp. Theor. Artif. Intell., vol. 35, no. 7, pp. 973–1000, Oct. 2022.

[29] H. Chen, Z. Wang, H. Jia, X. Zhou, and L. Abualigah, "Hybrid Slime Mold and Arithmetic Optimization Algorithm with Random Center Learning and Restart Mutation," Biomimetics, vol. 8, no. 5, p. 396, 2023.

[30] A. A. Ewees, F. H. Ismail, and A. T. Sahlol, "Gradient-based optimizer improved by Slime Mould Algorithm for global optimization and feature selection for diverse computation problems," Expert Syst. Appl., vol. 213, no. September 2022, 2023.

[31] A. A. Ewees et al., "Enhanced feature selection technique using slime mould algorithm: a case study on chemical data," Neural Comput. Appl., vol. 35, no. 4, pp. 3307–3324, 2023.

[32] S. Abid et al., "Development of Slime Mold Optimizer with Application for Tuning Cascaded PD-PI Controller to Enhance Frequency Stability in Power Systems," Mathematics, vol. 11, no. 8, 2023.

[33] L. Deng and S. Liu, "An enhanced slime mould algorithm based on adaptive grouping technique for global optimization," Expert Syst. Appl., vol. 222, no. February, p. 119877, 2023.

[34] A. K. Sharma, A. Saxena, and D. K. Palwalia, "Oppositional Slime Mould Algorithm: Development and application for designing demand side management controller," Expert Syst. Appl., vol. 214, no. January 2022, p. 119002, 2023.

[35] L. Wang, Q. Cao, Z. Zhang, S. Mirjalili, and W. Zhao, "Artificial rabbits optimization: A new bio-inspired meta-heuristic algorithm for solving engineering optimization problems," Eng. Appl. Artif. Intell., vol. 114, no. April, p. 105082, 2022.

[36] B. S. Yildiz, N. Pholdee, S. Bureerat, A. R. Yildiz, and S. M. Sait, "Enhanced grasshopper optimization algorithm using elite opposition-based learning for solving real-world engineering problems," Eng. Comput., vol. 38, no. 5, pp. 4207–4219, 2022.

[37] Y. Zhang, S. Du, and Q. Zhang, "Improved Slime Mold Algorithm with Dynamic Quantum Rotation Gate and Opposition-Based Learning for Global Optimization and Engineering Design Problems," Algorithms, vol. 15, no. 9, p. 317, Sep. 2022.

[38] B. S. Yıldız, S. Kumar, N. Pholdee, S. Bureerat, S. M. Sait, and A. R. Yildiz, "A new chaotic Lévy flight distribution optimization algorithm for solving constrained engineering problems," Expert Syst., vol. 39, no. 8, 2022.

[39] X. Yang et al., "An adaptive quadratic interpolation and rounding mechanism sine cosine algorithm with application to constrained engineering optimization problems," Expert Syst. Appl., vol. 213, no. PB, p. 119041, Mar. 2023.

[40] M. Abdel-Basset, R. Mohamed, M. Jameel, and M. Abouhawwash, "Nutcracker optimizer: A novel nature-inspired metaheuristic algorithm for global optimization and engineering design problems," Knowledge-Based Syst., vol. 262, p. 110248, Feb. 2023.

[41] F. S. Gharehchopogh, M. H. Nadimi-Shahraki, S. Barshandeh, B. Abdollahzadeh, and H. Zamani, "CQFFA: A Chaotic Quasi-oppositional Farmland Fertility Algorithm for Solving Engineering Optimization Problems," J. Bionic Eng., vol. 20, no. 1, pp. 158–183, Jan. 2023.

[42] L. Deng and S. Liu, "A multi-strategy improved slime mould algorithm for global optimization and engineering design problems," Comput. Methods Appl. Mech. Eng., vol. 404, p. 115764, 2023.

[43] A. Bala Krishna, S. Saxena, and V. K. Kamboj, hSMA-PS: a novel memetic approach for numerical and engineering design challenges, vol. 38, no. 4. Springer London, 2022.

[44] J. Zhao, Z. M. Gao, and H. F. Chen, "The Simplified Aquila Optimization Algorithm," IEEE Access, vol. 10, pp. 22487–22515, 2022.

[45] Z. M. Gao, J. Zhao, and S. R. Li, "The Improved Slime Mould Algorithm with Cosine Controlling Parameters," J. Phys. Conf. Ser., vol. 1631, no. 1, 2020.

[46] P. N. Suganthan et al., "Problem definitions and evaluation criteria for the CEC 2005 special session on real-parameter optimization," Tech. Report, Nanyang Technol. Univ. Singapore, May 2005 KanGAL Rep. 2005005, IIT Kanpur, India, no. May, pp. 1–50, 2005.

[47] S. Mirjalili, S. M. Mirjalili, and A. Hatamlou, "Multi-Verse Optimizer: a nature-inspired algorithm for global optimization," Neural Comput. Appl., vol. 27, no. 2, pp. 495–513, Feb. 2016.

[48] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," Adv. Eng. Softw., vol. 69, pp. 46–61, Mar. 2014.

[49] S. Mirjalili, "SCA: A Sine Cosine Algorithm for solving optimization problems," Knowledge-Based Syst., vol. 96, pp. 120–133, Mar. 2016.

[50] L. Abualigah, A. Diabat, S. Mirjalili, M. Abd Elaziz, and A. H. Gandomi, "The Arithmetic Optimization Algorithm," Comput. Methods Appl. Mech. Eng., vol. 376, p. 113609, 2021.

[51] S. García, A. Fernández, J. Luengo, and F. Herrera, "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power," Inf. Sci. (Ny)., vol. 180, no. 10, pp. 2044–2064, 2010.

APPENDIX A. ENGINEERING DESIGN PROBLEMS

A. Cantilever structure problem

$$\text{Minimize } f(x) = 0.6224(x_1 + x_2 + x_3 + x_4 + x_5)$$
$$\text{subject to, } g(x) = \frac{61}{x_1^3} + \frac{37}{x_2^3} + \frac{19}{x_3^3} + \frac{7}{x_4^3} + \frac{1}{x_5^3} - 1 \le 0$$
$$\text{Variable ranges: } 0.01 \le x_1, x_2, x_3, x_4, x_5 \le 100$$

B. The welded beam design problem

$$\text{Minimize } f(\vec{x}) = 1.10471x_1^2x_2 + 0.04811x_3x_4(14.0 + x_2)$$
subject to,
$$g_1(\vec{x}) = \tau(\vec{x}) - \tau_{max} \le 0$$
$$g_2(\vec{x}) = \sigma(\vec{x}) - \sigma_{max} \le 0$$
$$g_3(\vec{x}) = \delta(\vec{x}) - \delta_{max}$$
$$g_4(\vec{x}) = x_1 - x_4 \le 0$$
$$g_5(\vec{x}) = P - P_c(x) \le 0$$
$$g_6(\vec{x}) = 0.125 - x_1 \le 0$$
$$g_7(\vec{x}) = 0.1047x_1^2 + 0.04811x_3x_4(14.0 + x_2) - 5.0 \le 0$$
Variable ranges: $0.1 \le x_1 \le 2.0,\ 0.1 \le x_2 \le 10.0,\ 0.1 \le x_3 \le 10.0,\ 0.1 \le x_4 \le 2.0$
where,
$$\tau(x) = \sqrt{(\tau')^2 + 2\tau'\tau''\frac{x_2}{2R} + \tau''^2}, \tau' = \frac{P}{\sqrt{2}x_1x_2}, \tau'' = \frac{MR}{J}, M = P\left(L + \frac{x_2}{2}\right)$$
$$R = \sqrt{\frac{x_2^2}{4} + \left(\frac{x_1 + x_3}{2}\right)^2}, J = 2\left\{\sqrt{2}x_1x_2\left[\frac{x_2^2}{12} + \left(\frac{x_1 + x_3}{2}\right)^2\right]\right\}, \sigma(x) = \frac{6PL}{x_4x_3^2}, \delta(x) = \frac{4PL^3}{Ex_4x_3^3}$$
$$P_c(x) = \frac{4.103E\sqrt{\frac{x_3^2x_4^6}{36}}}{L^2}\left(1 - \frac{x_3}{2L}\sqrt{\frac{E}{4G}}\right), P = 6000lb, L = 14in, E = 30 \times 10^{06}psi$$
$$G = 12 \times 10^{06}psi. \tau_{max} = 136000psi, \sigma(x) = 30000psi, \delta_{max} = 0.25in$$

C. Pressure Vessel problem

$$\text{Minimize } f(\vec{x}) = 0.6224x_1x_3x_4 + 1.7781x_2x_3^2 + 3.1661x_1^2x_4 + 19.84x_1^2x_3$$
Subject to,
$$g_1(\vec{x}) = -x_1 + 0.0193x_3 \le 0,$$
$$g_2(\vec{x}) = -x_3 + 0.00954x_3 \le 0,$$
$$g_3(\vec{x}) = -\pi x_3^2x_4 - \frac{4}{3}\pi x_3^3 + 1{,}296{,}000 \le 0,$$
$$g_4(\vec{x}) = x_4 - 240 \le 0,$$
Variable ranges: $0 \le x_1 \le 99, 0 \le x_2 \le 99, 10 \le x_3 \le 200, 10 \le x_4 \le 200$

D. Compression Coil Spring design problem

$$\text{Minimize } f(\vec{x}) = (x_3 + 2)x_2x_1^2$$
subject to,
$$g_1(\vec{x}) = 1 - \frac{x_2^3x_3}{71785x_1^4} \le 0$$
$$g_2(\vec{x}) = \frac{4x_2^2 - x_1x_2}{12566(x_1^3x_2 - x_1^4)} + \frac{1}{5108x_1^2} - 1 \le 0$$
$$g_3(\vec{x}) = 1 - \frac{140.45x_1}{x_2^2x_3} \le 0$$
$$g_4(\vec{x}) = \frac{x_1 + x_2}{1.5} - 1 \le 0$$
Variable ranges: $0.05 \le x_1 \le 2.0, \quad 0.25 \le x_2 \le 1.3, \quad 2 \le x_3 \le 15.0$

E. Multiple disk clutch brake problem

$$\text{Minimize } f(\vec{x}) = \pi(x_2^2 - x_1^2)x_3(x_5 + 1)\rho,$$
Subject to,
$$g_1(x) = x_2 - x_1 - \Delta R \ge 0$$
$$g_2(x) = L_{max} - (x_5 + 1)(x_3 + \delta) \ge 0$$
$$g(x) = P_{max} - P_{rz} \ge 0$$
$$g(x) = P_{max} * Vsr_{max} - P_{rz} * Vsr \ge 0,$$
$$g_5(x) = Vsr_{max} - Vsr \ge 0,$$
$$g_6(x) = T_{max} - T \ge 0,$$
$$g_7(x) = M_h - sM_s \ge 0,$$
$$g_8(x) = T \ge 0,$$
Variable ranges: $60 \le x_1 \le 80, 90 \le x_2 \le 110, 1 \le x_3 \le 3, 0 \le x_4 \le 1000, 2 \le x_5 \le 9, i = 1,2,3,4,5.$
where,
$$M_h = \frac{2}{3}\mu x_4x_5\frac{x_2^3 - x_1^3}{x_2^2 - x_1^2}N.mm, W = \frac{\pi n}{30}rad/s, A = \pi(x_2^2 - x_1^2)mm^2$$
$$P_{rz} = \frac{x_4}{A}N/mm^2, Vsr = \frac{PiR_{sr}n}{30}mm/s, R_{sr} = \frac{2(x_2^3 - x_1^3)}{3(x_2^2x_1^2)}mm$$
$$\Delta R = 20mm, L_{max} = 30mm, \mu = 0.6, P_{max} = 1MP_a, p = 0.0000078\frac{kg}{mm^3}, Vsr_{max} = 10\frac{m}{s},$$
$$\delta = 0.5mm, s = 1.5, T_{max} = 15s, n = 250rpm, I_z = 55Kg.m^2, M_s = 40Nm, Mf = 3Nm$$

F. Speed reducer problem.

$$\text{Minimize } f(\vec{x}) = 0.7854x_1x_2^2(3.3333x_3^2 + 14.9334x_3 - 43.0934) - 1.508x_1(x_6^2 + x_7^2) + 7.4777(x_6^3 + x_7^3) + 0.7854(x_4x_6^2 + x_5x_7^2)$$
Subject to,
$$g_1(\vec{x}) = \frac{27}{x_1x_2^2x_3} - 1 \le 0,$$
$$g_2(\vec{x}) = \frac{397.5}{x_1x_2^2x_3^2} - 1 \le 0,$$
$$g_3(\vec{x}) = \frac{1.93x_4^3}{x_2x_3x_6^4} - 1 \le 0,$$
$$g_4(\vec{x}) = \frac{1.93x_5^3}{x_2x_3x_7^4} - 1 \le 0,$$

$$g_5(\vec{x}) = \frac{\sqrt{(\frac{745x_4}{x_2x_3})^2 + 16.9 \times 10^6}}{110x_6^3} - 1 \leq 0$$

$$g_6(\vec{x}) = \frac{\sqrt{(\frac{745x_5}{x_2x_3})^2 + 157.5 \times 10^6}}{85x_7^3} - 1 \leq 0$$

$$g_7(\vec{x}) = \frac{x_2x_3}{40} - 1 \leq 0,$$

$$g_8(\vec{x}) = \frac{5x_2}{x_1} - 1 \leq 0,$$

$$g_9(\vec{x}) = \frac{x_1}{12x_2} - 1 \leq 0,$$

$$g_{10}(\vec{x}) = \frac{1.5x_6 + 1.9}{12x_2} - 1 \leq 0,$$

$$g_{11}(\vec{x}) = \frac{1.1x_7 + 1.9}{x_5} - 1 \leq 0,$$

$$g_{11}(\vec{x}) = \frac{1.1x_7 + 1.9}{x_5} - 1 \leq 0,$$

Variable ranges: $2.6 \leq x_1 \leq 3.6, 0.7 \leq x_2 \leq 0.8, 17 \leq x_3 \leq 28,$ $7.3 \leq x_4 \leq 8.3,\ 7.8 \leq x_5 \leq 8.3, 2.9 \leq x_6 \leq 3.9, 5.5 \leq x_7 \leq 5$

G. Gear train engineering design problem

Minimize $f(\vec{x}) = \left(\frac{1}{6.931} - \frac{x_1x_2}{x_3x_4}\right)^2$

Variable ranges: $12 \leq x_1, x_2, x_3, x_4 \leq 60$

Research on Credit Card Fraud Prediction Model Based on GAN-DNN Imbalance Classification Algorithm

Qin Wang[1], Mary Jane C.Samonte[2*]
School of Information Technology, Mapua University,Xi'an Siyuan University, Manila, Phillippines[1]
School of Information Technology, Mapua University, Manila, Phillippines[2]

***Abstract*—Credit card consumption has become an important way of consumption in modern life, but the problem of credit card fraud has also emerged, disrupting the financial order and restricting the development of the industry. Aiming at the data class imbalance problem in credit card fraud detection and improving the accuracy of fraud detection, this paper uses the Generative Adversarial Network (GAN) to generate fraud samples and balance the number of fraud transaction samples and normal transaction samples. Then, a deep neural network (DNN) is used to construct a credit card fraud prediction model. The study compares this model with commonly used classification algorithms and sampling methods in detail and confirms that the designed credit card fraud prediction model has a good effect, providing a theoretical basis and practical reference for financial institutions to predict credit card fraud.**

Keywords*—Generative adversarial network; deep neural network; unbalanced data; credit card fraud; classification algorithms

I. INTRODUCTION

Globally, the field of credit card payments is rife with a significant number of credit card fraud cases. According to a report by the Federal Trade Commission (FTC), in 2022, as many as 2.8 million users in the United States were affected by credit card fraud, with losses amounting to $3.4 billion. By the end of 2022, the monetary losses caused by credit card fraud soared to $5.8 billion, an increase of over 70% [1]. In recent years, credit card fraud has shown a trend towards specialization, scaling, and syndication, undermining the stability of the financial market and endangering the financial security of the public [2]. Therefore, it is crucial for financial institutions to establish intelligent fraud detection mechanisms to enhance the anti-fraud capabilities of regulators.

In recent years, researchers have conducted extensive work in the area of credit card fraud detection. Regarding the issue of data imbalance, oversampling and undersampling methods are commonly employed. For instance, N. Rtayli et al. utilized the SMOTE algorithm to generate fraud samples to overcome data imbalance [3]; Chen Ying et al. improved the SMOTE algorithm through K-Means clustering, generating fraud samples only in safe areas [4]; E. Esenogho et al. combined SMOTE oversampling with undersampling for hybrid sampling, thereby improving the overall distribution of credit card data [5]. In terms of data detection accuracy, deep models based on Long Short-Term Memory (LSTM) networks are generally used [6]. Gao J et al., for example, used LSTM to extract potential temporal information from credit card data, ultimately completing information identification and fraud classification through XGBoost [7]; Benchaji et al. proposed an LSTM credit card fraud detection model integrated with an attention mechanism, selectively focusing on features through the attention mechanism to enhance the model's detection efficiency [8]; J. Forough et al. constructed a credit card fraud detection model with LSTM as the preliminary prediction layer and CRF as the final prediction layer [9]. However, the traditional SMOTE oversampling algorithm generates a large number of fraudulent samples with noise, and undersampling algorithms may lose key information, ultimately affecting the training effectiveness of the model [10]. On the other hand, LSTM can only learn the forward distribution of credit card data and cannot combine the forward and reverse directions to output a comprehensive expression for fraud detection [11]. In terms of data feature extraction and classification tasks, various hybrid model methods based on GAN have been adopted [12]. Li et al. used the GAN-RNN algorithm for sequence modeling of image data to mine time series features in images [13]; Zhang et al. combined GAN-RNN with variational autoencoders for feature fusion to enhance the expressive power of features [14]. Zhao et al. utilized GAN-CNN for in-depth mining of image features, ultimately completing image classification tasks through a Softmax classifier [15]; Sun et al. proposed a GAN-CNN classification model combined with a multi-scale feature fusion mechanism, improving the model's classification accuracy by fusing features of different scales [16]. However, GAN-RNN may lose local key information during feature extraction or fusion, affecting the model's comprehensive capture of features related to fraudulent behavior, thereby affecting the model's fraud detection capability [17]. GAN-CNN cannot adaptively adjust according to different transaction data features and fraud patterns, making it difficult to output more comprehensive and accurate fraud detection results [12].

Compared with traditional methods, this study employs a Generative Adversarial Network (GAN) to generate minority sample data and then combines it with a Deep Neural Network (DNN) to predict the categories of credit card transaction data. This approach has multiple advantages: it overcomes the issue of undersampling methods losing a large amount of data information, and also addresses the noise expansion and overfitting issues when the SMOTE oversampling method and

ADASYN oversampling method generate new samples. At the same time, applying the DNN model to the credit card fraud prediction problem expands the application range of deep learning technology and improves the model's predictive performance. In response to these issues, this paper proposes corresponding solutions. In terms of research methodology, it elaborates on the principles of Generative Adversarial Networks (GAN) and Deep Neural Networks (DNN), and constructs a credit card fraud prediction model. Through experimental design and analysis, including the introduction of the experimental environment, description and preprocessing of the dataset, and the determination of evaluation metrics, the experimental results are compared with common classification models and deep neural network classification models different data processing methods to further analyze model performance. Finally, the paper is summarized, highlighting the advantages of the research model and its reference value for financial institutions in credit card fraud prediction.

II. Research Methods

A. Generative Adversarial Networks (GAN)

Generative Adversarial Networks (GAN) [12] is an unsupervised neural network that consists of two networks for adversarial training. The goal of the generator is to synthesize samples that are difficult for the discriminator to distinguish, and the goal of the discriminator is to distinguish whether the samples generated by the generator are real samples as much as possible. In this paper, we use generative adversarial networks to generate fraud-like transaction data to solve the data imbalance problem in the original data. The modeling of the generative adversarial network is mainly divided into three aspects: the overall structure of the generative adversarial network, the modeling of the data generator and the modeling of the data discriminator.

- Overall structure of generative adversarial network

In this paper, Conditional Generative adversarial Network (CGAN), a derivative model of generative adversarial network, is used to generate fraudulent credit card transaction data to solve the problem of extremely unbalanced number of positive and negative samples in credit card transaction data. The conditional generative adversarial network model in the credit card fraud prediction model adds the credit card transaction data category label y to the data generator model and the data discriminator model in the model, so that the data generated by the generator is artificial credit card transaction data with the same distribution as the real data under the condition of meeting the category label y. This makes the adversarial network model easier to control, solves the uncertainty of the original generative adversarial network generation model, and makes the generated credit card transaction data more in line with our expectations. At the same time, it also makes the conditional generative adversarial network model can quickly reach the convergence conditions in the training process, the model is not easy to collapse, and the training process is easier to control. The overall structure of the conditional generative adversarial network in the credit card fraud prediction model is shown in Fig. 1.

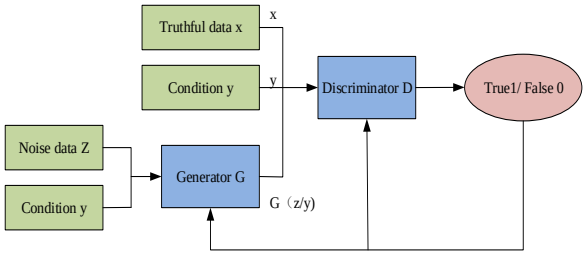


Fig. 1. Overall structure of conditional generative adversarial networks for credit card fraud prediction.

The objective function of the Conditional Generative Adversarial Network (CGAN) model in the credit card fraud prediction model is expressed in Eq. (1). In this equation, pdate represents the distribution of real credit card transaction data, y represents the category label of credit card transactions, pz represents the distribution of credit card transaction data generated by the generator, and y' represents the category label generated by the generator for the credit card transactions. When training the generator, the goal is to maximize V(D, G), whereas when training the discriminator, the goal is to minimize V(D, G).

$$\min_G \max_D V(D,G) = E_{x\sim P_{date}(x)}\left[\log D(x|y\right] + E_{z\sim p_z(z)}\left[\log(1-D(G(x|y)))\right] \quad (1)$$

- Data generator model

Credit card transaction data has the characteristics of huge data volume and moderate data feature dimension, and good results can be achieved by using a fully connected deep neural network. Therefore, the data generator model of the conditional generative adversarial network in the credit card fraud prediction model in this study adopts a fully connected deep neural network. In this model, the random noise data z conforming to Gaussian distribution and the credit card transaction data category label y are combined as the input of the data generator model, and the input layer and the hidden layer are connected in a fully connected manner. In addition, in order to improve the fitting ability of the deep neural network, a nonlinear activation function needs to be added between the layers. However, the sigmoid activation function is easy to cause the problem of gradient disappearance. Therefore, leakyrelu activation function needs to be added between the input layer and the hidden layer in the data generator model. Each hidden layer is also connected with the upper hidden layer in a fully connected way, and the leakyrelu activation function is used, and the connection between the last layer and the output layer is also fully connected. The network structure of the data generator model in the credit card fraud prediction model is shown in Fig. 2.

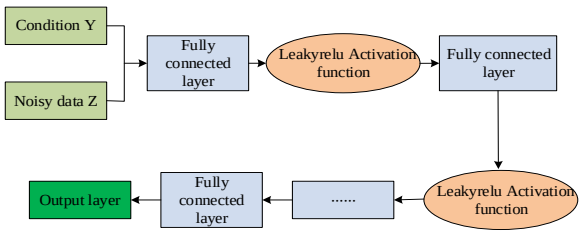


Fig. 2. Network structure of data generator model.

The training of the data generator model should ensure that the data discriminator model does not change, and the goal of the data generator model is to make the data discriminator not be able to determine whether the data is real credit card transaction data or credit card transaction data generated by the generator. Therefore, the loss function of the data generator model is shown in Eq. (2):

$$V(G) = E_{z\sim p_z(z)}\left[\log(1-D(G(z|y)))\right] \quad (2)$$

The training process of the data generator model is the process that minimizes this V(G), the data generator model Algorithm 1 shows the training process.

Algorithm 1: Training process of data generator model

Input：Random noise data z, label y

Output：Data generator

（1）Initializes the data generator

（2）The sample set is composed of m minibatch samples collected from pg(z) of Gaussian distribution

$$Z = \left\{z^{(1)}, z^{(2)}, z^{(3)}, \cdots, z^{(m)}\right\}$$

（3）The sample set of noise is connected to the class label y to obtain the input of the data generator

$$input = \left\{(z^{(1)}, y), (z^{(2)}, y), (z^{(3)}, y), \cdots (z^{(m)}, y)\right\}$$

（4）The input data input passes through multiple hidden layers and the leakyrelu activation function to get the output data

$$output = \left\{z_g^{(1)}, z_g^{(2)}, z_g^{(3)}, \cdots, z_g^{(m)}\right\}$$

（5）The gradient descent algorithm is used to update the parameters of the data generator model to make the loss function of the generator

$$V(G) = E_{z\sim p_z(z)}\left[\log(1-D(G(z|y)))\right]$$

Min.

- Data discriminator model

In addition, in the credit card fraud prediction model, the data discriminator model is required to correctly identify fraud transaction samples when solving the problem of data class imbalance. Therefore, adding the category label y to the input of the data discriminator can make the training process of the data discriminator model more targeted. The network structure of the data discriminator model in the credit card fraud prediction model is shown in Fig. 3.

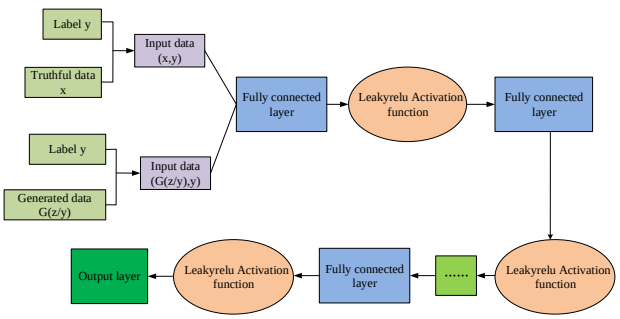


Fig. 3. Network structure of data discriminator model.

While training the data discriminator model, we need to ensure that the data generator model does not change, and the goal of the data discriminator is to make the discriminator maximally distinguish between the credit card transaction data generated by the generator and the real credit card transaction data. Therefore, the loss function of the data discriminator model is shown in Eq. (3):

$$V(D) = E_{x\sim p_{date}(x)}\left[\log D(x|y)\right] \quad (3)$$

The learning purpose of the data discriminator model in the conditional generation adversarial network is to maximize V(D). The specific training process is shown in Algorithm 2.

Algorithm 2: Training process of data discriminator model

Input：Data generator generated data G(z|y), real data x, label y

Output：Data discriminator

（1）Initializes the data discriminator

（2）From the distribution $p_{date}(x)$ of the real credit card transaction data, m minibatches of data are randomly selected to form the data set of the real data

（3）From the data G(z|y) sample generated by the data generator, m minibatch data are randomly selected to form the data set generated by the generator

$$G = \left\{g^{(1)}, g^{(2)}, g^{(3)}, \cdots, g^{(m)}\right\}$$

（4）A dataset of real data is concatenated with label y to obtain a portion of the discriminator input.

$$input1 = \left\{(x^{(1)}, y), (x^{(2)}, y), (x^{(3)}, y), \cdots, (x^{(m)}, y)\right\}$$

（5）The data set of generated data is concatenated with label y to obtain a portion of the discriminator input

$$input2 = \left\{(g^{(1)}, y), (g^{(2)}, y), (g^{(3)}, y), \cdots, (g^{(m)}, y)\right\}$$

（6）input1 and input2 are mixed to get the discriminator input

$$input = \left\{(x^{(1)}, y), (x^{(2)}, y), \cdots, (x^{(m)}, y), (g^{(1)}, y), (g^{(2)}, y), \cdots, (g^{(m)}, y)\right\}$$

（7）The input data input passes through the hidden layer and the leakyrelu activation function to get the output data

$$logits = \left\{z_g^{(1)}, z_g^{(2)}, z_g^{(3)}, \cdots, z_g^{(m)}\right\}$$

（8）The output logits of the last layer of the hidden layer is used by the sigmoid activation function to obtain the output of the entire discriminator

$$\text{output} = sigmoid(\log its)$$

（9）The gradient ascending algorithm is used to update the parameters of the discriminator and make the loss function of the discriminator

$$V(D) = E_{x\sim p_{date}(x)}\left[\log D(x|y)\right]$$

Max.

B. Deep Neural Networks

The neural network model has a strong fitting ability, and the deep neural network is deeper than the ordinary neural network, that is, the hidden layers are more, and its fitting ability is relatively stronger. The deep neural network model is shown in Fig. 4.

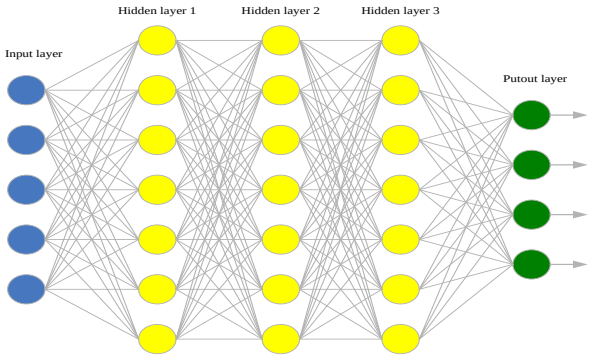


Fig. 4. Deep neural networks model.

According to the characteristics of credit card historical transaction data, the deep neural network classifier model of credit card fraud prediction model can obtain good results by using the deep fully connected neural network model. The network structure of the deep neural network classifier in the credit card fraud prediction model is shown in Fig. 5.

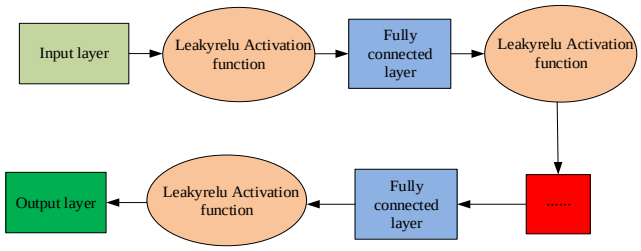


Fig. 5. Network structure of deep neural network classifier.

In the training process of deep neural network classifier, cross entropy can be used as a loss function. The definition of cross entropy is shown in Eq. (4):

$$Loss = \frac{1}{N}\sum_i -\left[y_i \cdot \log(p_{i)} + (1-y_i)\cdot \log(1-p_i)\right] \quad (4)$$

In the cross entropy loss function, N represents the number of samples, yi represents the actual value of the i th sample, the fraudulent transaction sample is 1, the normal transaction sample is 0, pi represents the predicted value of the i th sample, that is, the probability value of the i sample predicted as the fraudulent sample. Cross entropy loss function is often used in conjunction with softmax activation function, which can improve the speed and effectiveness of model training to a certain extent. The detailed training process of deep neural network classifier is shown in Algorithm 3.

Algorithm 3: Training process of deep neural network classifier

Input：Training set

$X = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)}), \cdots, (x^{(m)}, y^{(m)})\}$, among $y^i = \{0,1\}$

Output：Deep neural network classifier D

（1）Initialize all variables in a deep neural network, including weight vectors and paranoid value vectors

（2）Loop iteration □ = 1,2,3,⋯,□：

①Calculate the output of the classifier based on the current parameters and formulas

②The parameters of the classifier are updated by gradient descent algorithm to make the cross entropy loss function

$$Loss = \frac{1}{N}\sum_i -\left[y_i \cdot \log(p_{i)} + (1-y_i)\cdot \log(1-p_i)\right]$$

Min.

C. Credit Card Fraud Prediction Model

The credit card fraud prediction model uses generative adversarial networks to deal with the class imbalance problem in credit card transaction data, and then combines the deep neural network model in the field of deep learning to predict the class of credit card transaction samples. An example plot of a credit card fraud prediction model is shown in Fig. 6.

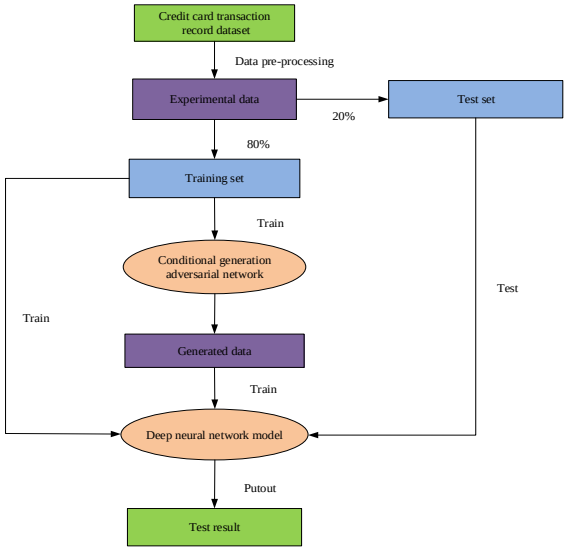


Fig. 6. Credit card fraud prediction model.

In the credit card fraud prediction model of this study, the generative adversarial network is used to generate fraud sample data to solve the extreme imbalance problem between fraud sample data and normal transaction sample data in the original credit card transaction data set, and then the high fitting ability of deep neural network is used to predict the credit card transaction category.

III. Experimental Design and Analysis

A. Experimental Environment

This article was tested using Windows 11, AMD Ryzen 7 5800H CPU, 16 GB of memory, NVIDIA GeForce RTX 3060 display card. In the process of building the model, Sklearn toolkit and Keras toolkit of Python programming language are used to achieve.

B. Data Set Description and Preprocessing

The dataset was collected from November 2022, through a professional consulting company, including open data information from volunteers, commercial banks, private banks, and credit cooperatives. Currently, the dataset contains transaction records of 10,000 credit card users, including different regions, age groups, and professional backgrounds, ensuring the diversity and representativeness of the data. Among the 10,000 users, after strict data cleaning and labeling, it is confirmed that 178 users have fraudulent behavior, that is, the proportion of fraudulent users is 1.78%. Although this proportion is not high, considering the potential harm of fraudulent behavior to financial institutions, accurate prediction is particularly important.

There are 31 feature variables in this dataset with no missing values. For the purpose of protecting cardholder privacy, their original characteristics and background information are not provided. Features V1, V2, ... V28 is the result of dimensionality reduction through PCA. The specific information of the data set is shown in Table I:

TABLE I. DATASET ATTRIBUTE TABLE

Ordinal	Field Name	Data Type	Field Description
1	Time	Float	The number of seconds elapsed between this transaction and the first transaction in the dataset
2-29	V1-V28	Float	Principal component data
30	Amount	Float	Transaction amount
31	Class	Float	In the case of fraud, take it as one, otherwise take zero.

C. Credit Card Fraud Prediction Model

While constructing and training the model, we need to establish a rational and effective indicator system to verify and evaluate the model's performance. The evaluation indicators used in this paper include Accuracy, Precision, Recall calculation formula is as follows:

$$Accurary = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$\Pr ecision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

Among, TP represents the sample that predicted a positive example as a positive example; FP is the sample that predicts a negative example as a positive example. FN is denoted as the sample that predicts a positive example as a negative example; TN denotes the sample for which a negative example is predicted as a negative example.

IV. ANALYSIS OF EXPERIMENTAL RESULTS

In order to verify the effectiveness of the model proposed in this study, we use credit card transaction data set to train common classification models and deep neural network classification models for detailed comparison. The performance effect is shown in Table II.

TABLE II. CLASSIFIER TEST RESULTS

Model	accuracy	precision	recall
Logistic regression [18]	0.8623	0.9836	0.8612
Naive Bayes [19]	0.7657	0.9829	0.7634
Decision-Making tree [20]	0.8526	0.9882	0.8510
Kproximity Classifier [21]	0.7584	0.9800	0.7534
SVM [22]	0.8321	0.9902	0.8308
AdaBoost classifier [23]	0.8112	0.9938	0.8094
Dnn classifier [24]	0.9232	0.9985	0.9192

In order to verify the effectiveness of conditional generative adversarial networks in generating new samples, In this study, the experimental results of combining conditional generative adversarial network model with deep neural network classifier and the experimental results of not processing imbalanced data only using deep neural network classifier, the experimental results of combining random undersampling with deep neural network classifier, the experimental results of combining SMOTE oversampling with deep neural network classifier, and the experimental results of combining SMOTE oversampling with deep neural network classifier The test results of ANASYN oversampling combined with deep neural network classifier were compared, and the final results are shown in Table III.

TABLE III. EXPERIMENTAL RESULTS OF VARIOUS UNBALANCED DATA PROCESSING METHODS

Model	Accuracy	Precision	Recall
DNN [25]	0.9232	0.9985	0.9192
Random undersampling+dnn	0.9280	0.9987	0.9348
SMOTE+dnn	0.9250	0.9983	0.9192
ANASYN+dnn	0.9300	0.9989	0.9406
CGAN+dnn	0.9350	0.9991	0.9501

The ROC curve is shown as follows (Fig. 7):

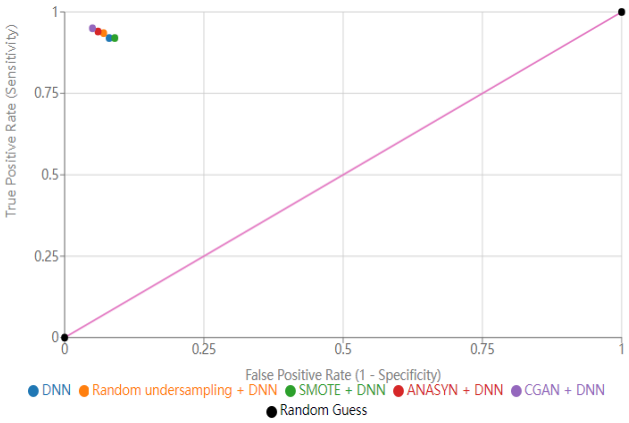


Fig. 7. ROC curve analysis chart.

Fig. 6 presents a performance comparison of various deep neural network (DNN) variants in the receiver operating characteristic (ROC) space. These variants include the basic DNN model as well as DNN models combined with different data balancing techniques such as random undersampling, Synthetic Minority Over-sampling Technique (SMOTE), Analytic Synthetic (ANASYN), and Conditional Generative Adversarial Network (CGAN). By analyzing the distribution of these models in the ROC space, we can draw several key observations, among which the most remarkable is the outstanding performance of the CGAN+DNN model.

The CGAN+DNN model stands out among all tested models and shows the optimal performance indicators. This model achieves a true positive rate (TPR) of 0.95 and a false positive rate (FPR) of only 0.05. This result is closest to the ideal upper left corner position in the ROC space. The excellent performance of CGAN+DNN can be attributed to the unique advantage of CGAN in generating high-quality synthetic samples. By generating realistic and diverse minority class samples, CGAN effectively alleviates the data imbalance problem, thereby significantly improving the classification performance of DNN.

The advantages of the CGAN+DNN model are more prominent in comparison with other models:

Compared with the basic DNN model, CGAN+DNN has achieved significant improvements in both TPR and FPR indicators. This indicates that CGAN not only improves the model's ability to recognize positive examples but also reduces the probability of misjudging negative examples as positive examples.

Compared with SMOTE+DNN, CGAN+DNN shows obvious advantages. As a traditional oversampling technique, SMOTE seems to fail to effectively improve the performance of DNN in this study. This comparison highlights the superiority of CGAN in generating complex and high-quality synthetic samples.

Although the ANASYN+DNN model also performs very well (TPR is 0.94 and FPR is 0.06), CGAN+DNN is still slightly better. This small but significant difference reflects the superiority of CGAN in capturing the complexity of data distribution.

Random undersampling+DNN performs quite well in this study, but still cannot surpass CGAN+DNN. This comparison not only emphasizes the advantages of CGAN but also reveals that in some cases simple techniques may also produce good results.

The outstanding performance of the CGAN+DNN model is not only reflected in numbers but more importantly in its potential in practical applications. When dealing with highly imbalanced datasets, CGAN can generate diverse and real synthetic samples, which is particularly important in key fields such as medical diagnosis and fraud detection. The samples generated by CGAN can help the model learn richer feature representations, thereby effectively controlling FPR while maintaining a high TPR, which is crucial in many practical applications.

V. Conclusion

This paper proposes an innovative method to address the challenges of credit card fraud detection. By leveraging generative adversarial networks (GAN) to generate synthetic fraud samples, the problem of class imbalance in credit card transaction data is alleviated. The authors present a deep neural network (DNN) model that, in combination with the samples generated by GAN, predicts credit card fraud with high accuracy. The research also conducts a comprehensive comparison with existing classification algorithms and sampling methods, demonstrating the effectiveness of the proposed model in enhancing the ability of financial institutions to predict and prevent fraud.

The innovation points of this article: The use of conditional generative adversarial networks (CGAN) to generate synthetic fraud samples, which introduces a new dimension in dealing with the class imbalance problem by incorporating class labels into the model. The combination of CGAN and DNN creates a powerful prediction model that not only addresses the data imbalance issue but also utilizes the strong fitting ability of DNN. The model is capable of generating high-quality synthetic samples, facilitating the training process, thereby increasing the detection rate and achieving more controlled convergence during model training.

Facing industry constraints and challenges in data collection, our current research is predicated on a relatively modestly sized dataset. In the future, we have already planned to employ a more extensive dataset, which will be reflected in our subsequent studies. We have already charted a course that includes the utilization of a plethora of machine learning models for broader and more profound testing, thereby enhancing the robustness and scalability of our models. This will further validate and refine the model's performance across a wider and more diverse array of credit card transactions. We will also explore an expanded repertoire of machine learning models and their integration with Generative Adversarial Networks (GANs) to bolster the model's resilience and scalability. Additionally, we will investigate the model's long-term predictive accuracy and its capacity to adapt to the ever-evolving patterns and techniques of fraud. Based on the proposed model, we aim to develop a real-time fraud detection system that provides immediate alerts to financial institutions.

References

[1] FEDERAL TRADE COMMISSION. Consumer Sentinel Network Data Book 2021 [EB/OL] (2022-02-22) [2022-03-22].https : // www. ftc. gov/reports/consumer-sentinel-network-data-book-2021.

[2] CARCILLO F, BORGNE Y, CAELEN O. Combining Unsupervised and Supervised Learning in Credit Card Fraud Detection [J] .Information Sciences, 2019, 557 : 317-331.

[3] RTAYLI N, ENNEYA N. Enhanced credit card fraud detection based on SVM recursive feature elimination and hyper parameters optimization [J]. Journal of information security and applications, 2020, 55: 102596.

[4] CHEN Ying, ZHANG Ruirui. Research on credit card default prediction based on K Means SMOTE and BP neural network [J]. Complexity, 2021: 6618841.

[5] ESENOGHO E, MIENYE I D, SWART T G, et al. A neural network ensemble with feature engineering for improved credit card fraud detection [J]. IEEE access, 2022, 10: 16400-16407.

[6] HOCHREITER S, SCHMIDHUBER J. Long short term memory [J]. Neural computation, 1997, 9(8): 1735-1780.

[7] GAO J, SUN W J, SUI X.credit card users based on XGBoost LSTM model Research on default prediction for [J]. Discrete dynamics in nature and society, 2021(2): 5080472.

[8] BENCHAJI I, DOUZI S, EL OUAHIDI B, et al. Enhanced credit card fraud detection based on attention mechanism and LSTM deep model [J]. Journal of big data, 2021, 8(1): 151.

[9] FOROUGH J, MOMTAZI S. Sequential credit card fraud detection: A joint deep neural network and probabilistic graphical model approach [J]. Expert systems, 2022, 39(1):e12795.

[10] Lei Cheng, Xiaofu Wu, Suofei Zhang. Analysis of the Influence of Dataset Class Imbalance on Transfer Learning[J]. Signal Processing, 2020, 36(1): 110-117. DOI: 10.16798/j.issn.1003-0530.2020.01.014.

[11] Vikash Chander Maheshwari, Nurul Aida Osman, Norshakirah Aziz. A Hybrid Approach for Credit Card Fraud Detection Using Deep Neural Networks and Attention Mechanism. Journal of Advanced Research in Applied Sciences and Engineering Technology, 2023, 32(1): 315331. DOI: 10.37934/araset.32.1.315331.

[12] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets [C]// Conference on Neural Information Processing Systems. Montreal, Canada: ACM, 2014: 2672.

[13] Li Y, Chen X, Hwang T. A Financial Fraud Detection Model Based on LSTM Deep Learning Technique [J]. Journal of Applied Security Research, 2019, 14(1): 76-90. DOI: 10.1080/19361610.2019.1579422.

[14] Zhang H, Xu T, Li H, et al. StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(8): 1947-1962.

[15] Zhao W, Xu Y, Chen A, et al. Global Filter Networks for Image Classification [C]// International Conference on Neural Information Processing Systems. Montreal, Canada: ACM, 2021: 2672. DOI: 10.1145/3456789.3456790.

[16] Sun W, et al. Multi-feature fusion: Graph neural network and CNN combining for hyperspectral image classification [J]. Information Sciences, 2022, 580: 241-255. DOI: 10.1016/j.ins.2022.04.081.

[17] Zhao W, Xu Y, Chen A, et al. Global Filter Networks for Image Classification [C]// International Conference on Neural Information Processing Systems. Montreal, Canada: ACM, 2021: 2672.

[18] Brownlee J. Logistic Regression for Credit Card Fraud Detection [C]// International Conference on Machine Learning and Data Mining. Sydney, Australia: Springer, 2019: 123 - 135.

[19] Zhang L, Wang Y. Naive Bayes Approach in Credit Card Fraud Detection [C]// International Symposium on Computational Intelligence and Applications. Tokyo, Japan: IEEE, 2020: 45 - 52.

[20] Li X, Liu Z. Decision - Making Tree Algorithm for Credit Card Fraud Prediction [C]// Asian Conference on Artificial Intelligence. Seoul, South Korea: ACM, 2018: 302 - 310.

[21] Chen H, Ma Q. K - proximity Classifier in Credit Card Fraud Detection [C]// International Conference on Data Science and Big Data Analytics. Dubai, United Arab Emirates: Elsevier, 2022: 278 - 286.

[22] Wang S, Zhang Y. Support Vector Machine for Credit Card Fraud Detection [C]// European Conference on Machine Learning. Paris, France: Springer, 2021: 87 - 95.

[23] Zhao J, Hu M. AdaBoost Classifier in Credit Card Fraud Detection [C]// Global Conference on Artificial Intelligence and Machine Learning. London, United Kingdom: IEEE, 2023: 108 - 116.

[24] Liu G, Li T. Dnn Classifier for Credit Card Fraud Prediction [C]// World Congress on Computational Intelligence. Honolulu, United States: ACM, 2022: 182 - 190.

[25] Smith A, Johnson B. Deep Neural Network for Credit Card Fraud Detection [C]// International Conference on Deep Learning and Applications. San Francisco, United States: Springer, 2023: 456 - 468.

Impact Analysis of Informatization Means Driven by Artificial Intelligence Technology on Visual Communication

Lei Ni
School of Art and Design, Yellow River Conservancy Technical Institute, Kaifeng 475004, China

***Abstract*—With the popularization of computer technology, the combination of artificial intelligence and image processing technology has become a research hotspot in the visual communication. Image processing technology mostly involves segmentation and detection of images. Image segmentation often focuses on extracting image contour information, while ignoring the color of the image. The calculation time for image detection is relatively long, and the calculation steps are also relatively cumbersome. In response to the above issues, a density peak clustering algorithm was proposed for image segmentation. In the phase of image detection, the region recommendation network is introduced to improve the faster region Convolutional neural network algorithm. The findings demonstrate that under 15% Gaussian noise and 10% Salt-and-pepper noise, the segmentation accuracy of the density peak clustering algorithm is 98.13% and 97.89% respectively. The accuracy, recall and F-measure of the improved fast region Convolutional neural network algorithm are 98.49%, 97.29% and 97.77% respectively. The accuracy and average time consumption in the graphics processor environment are 98.18% and 2.94ms, respectively. In conclusion, the image segmentation algorithm based on density peak clustering algorithm and the improved fast region Convolutional neural network algorithm are robust, which have good segmentation and detection effects.**

***Keywords*—Image segmentation; image detection; density peak clustering algorithm; convolutional neural network; faster region convolutional neural network**

I. Introduction

A. Research Background

With the continuous progress of science and technology,the Visual communication is also seeking breakthroughs [1-2]. As a common form of visual communication, the processing technology of images has also shifted from artificial to intelligent [3]. Especially in today's increasingly popular digital information processing and transmission, images, as the most intuitive form of visual expression, have become an indispensable part of information communication. Image processing technology includes image digitization, image enhancement and restoration, image data encoding, image segmentation (IS), and image detection technology. Covering various aspects from image acquisition, image analysis to image understanding, the progress of these technologies directly promotes the rapid development of application fields such as medical imaging, autonomous driving, intelligent security, entertainment media, etc. [4]. As a key link in image processing technology, image segmentation and detection are becoming increasingly important. Among them, IS technology divides an image into several specific and unique regions. Then the area is extracted and displayed independently [5-7]. The application of this technology can help computers identify and extract regions of interest in images, while image detection technology determines their categories and locations through further analysis and localization of these regions. Image detection technology refers to IS based on target geometry and statistical features. Furthermore, the segmentation and recognition of the target are combined to determine the location and size of the target [8]. In recent years, with the introduction of deep learning algorithms, especially the successful application of convolutional neural networks (CNN) in image processing, the accuracy and speed of image segmentation and detection have been greatly improved. Meanwhile, emerging technologies such as convolutional neural networks combined with region recommendation networks also provide broad space for further optimizing image processing.

B. Existing Issues

Although deep learning has made significant progress in image segmentation and detection, there are still some issues that need to be addressed urgently. First, traditional IS algorithms often use image contour information to extract regions of interest in images. The importance of color information in images is ignored, resulting in computational redundancy. Traditional image detection algorithms are mostly based on the depth CNN to detect image objects. Its computation takes a long time, which can easily lead to computational redundancy and a large amount of memory occupation, resulting in slow image object detection speed. Secondly, the widely used convolutional neural networks have a large computational load and slow detection speed when processing image detection. Due to the complexity of convolution operations, especially when generating candidate regions, a large number of redundant computational steps can lead to excessive memory consumption and significantly increased processing time, making it difficult to meet real-time requirements [9]. In addition, existing algorithms perform poorly in processing images with high noise or complex backgrounds, easily losing key details and leading to a decrease in detection accuracy. Finally, some existing image segmentation and detection methods mostly rely on the specific distribution of data, and the algorithm has poor universality to the dataset. The robustness and adaptability of algorithms still need to be improved when facing different application scenarios or data types. Especially in practical applications, the

*Corresponding Author.

diversity and complexity of image data pose higher requirements for algorithms. Therefore, designing an algorithm that can simultaneously handle multiple complex situations, improve segmentation and detection accuracy, and has efficient computing capabilities has become an important research direction in the field of visual communication.

C. Research Content and Innovative Points

In response to the above issues, based on the low requirement for data distribution and simple principle of the Density Peak Clustering (DPC), the image center information and color information are extracted separately for IS work to obtain the DPC-IS algorithm. In order to achieve the image detection target, the fast-regional Convolutional neural network (FastR-CNN) algorithm is improved by introducing the regional recommendation network. The innovation points of the research mainly include the following two points. Firstly, DPC-IS algorithm is used to extract the center information and color information, which is convenient for Hierarchical clustering of pixels in the image. Then, based on the FastR-CNN algorithm, a regional recommendation network is introduced to reduce computational redundancy and improve detection accuracy and speed. The research structure contains four parts. The first is a review of relevant research results. The second is the DPC-IS algorithm and the FasterR-CNN algorithm construction. The third part is to verify the availability of the IS and image detection. The final part is a summary.

II. Related Work

IS is an important image-processing technology. The purpose is to segment the image into practical and non-interference regions. Objects of interest in the area can be displayed independently from the Background independence. As a hotspot direction in computer vision, how to apply IS technology to different professional fields has been deeply explored by many scholars. Shi et al. designed a fuzzy contour model based on multi-channel CNN to enhance the segmentation accuracy of medical images. The findings demonstrate that the model has good segmentation performance and robustness [10]. Liu et al. designed an IS algorithm based on deep learning network to identify the particle size distribution of crushed ore. Firstly, this algorithm was applied to preprocess the raw ore images. Then, deep learning networks were used to generate optimization model. The feasibility of this algorithm has been verified through simulation experiments [11]. Ji proposed a Level set algorithm based on significant fitting energy to solve the noise and non-uniformity of gray levels in image segmentation. This algorithm introduced a distance regularization term to eliminate the reinitialization process. Simulation experiments show that the algorithm has good segmentation ability in visual perception [12]. To improve the noise resistance and discrimination of segmented images with uneven grayscale, Li et al. designed a Bayesian criterion based on image segmentation. Local prior region descriptions were used to characterize image regions. From the simulation experiment, this method has superior performance in time efficiency and noise robustness [13]. Biswas et al. proposed a segmentation image active contour model to improve the image segmentation effect by combining the local binary fitting energy function and the improved Laplacian Gaussian function energy function. The findings indicate that the active contour model has good performance in segmentation accuracy, F-value, and Central Processing Unit (CPU) execution time [14].

Image detection technology is another important means in image processing technology. The purpose is to enable computers to automatically recognize target categories in images. By drawing a bounding box around the target, the position of each target is labeled. Chen proposed a license plate recognition model based on CNN and K-means clustering segmentation to address the recognition difficulties caused by uneven license plate form, color, viewpoint, and lighting environment. From the findings, this model can improve the recognition and detection accuracy [15]. Vinolin et al. proposed a deep CNN based on Taylor rider optimization algorithm for detecting concatenated images. Compared with traditional detection algorithms, precision measurement accuracy has been significantly improved [16]. Based on the fast CNN model, Luo et al. introduced neural architecture search optimization and feature enrichment to achieve multi-scale vehicle target detection in traffic scenes. From the findings, the robustness of the detection method is effectively improved [17]. Ramaraj proposed a face image detection algorithm based on CNN with Long short-term memory, aiming at the difficulty of face image detection caused by facial occlusion and low resolution. The research results confirm that the detection algorithm has superior performance and good detection accuracy [18]. Joy et al. constructed a fast-regional CNN multi object detection method based on incremental classification in the monitoring field. This detection method was mainly divided into class incremental learning part and domain adaptive part. The experimental results confirm that this method performs well on challenging targets including lighting changes, shadows, partial occlusion, and dynamic backgrounds [19].

To summarize, in image processing technology, there are rich research achievements in IS algorithm and image detection algorithm. But most IS algorithms only utilize the contour information of the image, ignoring the importance of the color information of the image. Most image detection algorithms require a significant amount of time when generating candidate regions through selective search. This process will generate a large number of invalid regions, resulting in wasted computing power. To address the above issues, a DPC-IS algorithm is constructed for IS based on spatial color distribution. The FasterR-CNN algorithm is constructed to detect target images.

III. DPC-IS Algorithm Design and FasterR-CNN Algorithm Construction

As a visual communication medium, images can express information in the most intuitive and authentic way. Therefore, compared to other media, images convey a richer amount of information. To process images more intuitively, DPC-IS algorithm and FasterR-CNN algorithm are used for segmentation and detection. This chapter focuses on the DPC-IS algorithm design and the FasterR-CNN algorithm construction in image detection.

A. Design of DPC-IS Algorithm in Image Segmentation

In Visual communication, image is not only a widely used form, but also a unique data format. Pixels are evenly arranged in the image space. Each pixel carries a unique color label. And each pixel has two features, namely spatial position feature (X,Y) and color feature (R,G,B). The spatial and color features are uniformly set to 1 dimension. The spatial and color features are merged to form a two-dimensional space. A two-dimensional sparse matrix (i,j) is used to represent an image. The value of matrix (i,j) is 0 or 1, indicating whether the pixel value of image i is j, as shown in Fig. 1.

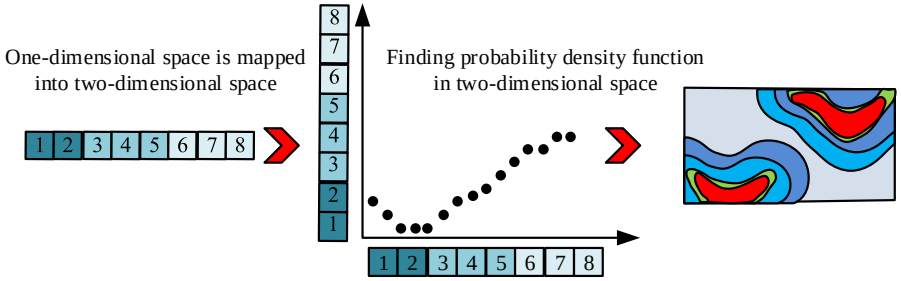


Fig. 1. The 1-dimensional space is mapped to the 2-dimensional space and the distribution of probability density function is obtained.

After obtaining the two-dimensional spatial distribution position of image pixels in Fig. 1, the kernel probability density function is used to measure the distribution of pixel points. For a set $\{X_1, X_2, \cdots, X_n\}$ with n sample points, the probability density function $f_g(x)$ of a certain point is equal to the sum of the influences of all data samples at that point, as shown in Eq. (1)

$$f_g(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X - X_i}{h}\right) \tag{1}$$

In Eq. (1), n is the sample points. K represent the kernel function. h is the window width. When using kernel functions to estimate the probability density function, the distance $d\left(\left|P_i, P_j\right|\right)$ between data points P_i and P_j in the 5-dimensional space (X,Y,R,G,B) needs to be defined, as shown in Eq. (2).

$$d\left(\left|P_i, P_j\right|\right) = \sqrt{\left(\frac{d_s}{\alpha_s}\right)^2 + \left(\frac{d_c}{\alpha_c}\right)^2} \tag{2}$$

In Eq. (2), d_s and d_c represent spatial distance and color distance, respectively. α_s and α_c are spatial smoothing parameters and color smoothing parameters, respectively. The values of α_s and α_c need to be manually set. The expression of spatial distance d_s is shown in Eq. (3).

$$d_s = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \tag{3}$$

In Eq. (3), x_i and x_j represent the horizontal distances of i and j. y_i and y_j stands for the vertical distances of i and j, respectively. The color distance d_c is expressed as an approximate CIElab color distance, as shown in Eq. (4).

$$dc = \sqrt{\left(2 + \frac{\bar{r}}{256}\right) \times \Delta R^2 + 4 \times \Delta G^2 + \left(2 + \frac{255 - \bar{r}}{256}\right) \times \Delta B^2} \tag{4}$$

In Eq. (4), ΔR, ΔG and ΔB stands for the weighted Euclidean distances of red, green, and blue, respectively. $\bar{r}$ refers to the average level of red. The expressions of ΔR, ΔG and ΔB are shown in Eq. (5).

$$\begin{cases} \Delta R = r_i - r_j \\ \Delta G = g_i - g_j \\ \Delta B = b_i - b_j \end{cases} \tag{5}$$

In Eq. (5), r_i and r_i represent the red distances of i and j. g_i and g_j stands for the green distances of i and j. b_i and b_j represent the blue distances of i and j, respectively. The probability density function f of the kernel function is shown in Eq. (6).

$$f = \frac{1}{n}\sum \exp\left(\frac{d_s^{\,2}}{\alpha_s^{\,2}} + \frac{d_c^{\,2}}{\alpha_c^{\,2}}\right) \tag{6}$$

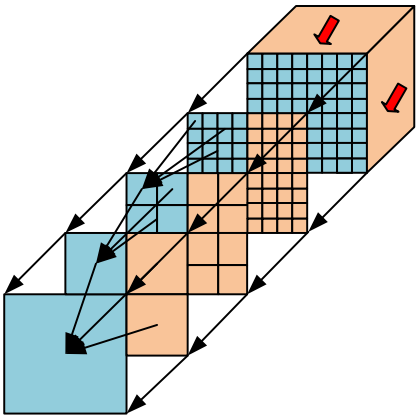


Fig. 2. Schematic diagram of the reduction algorithm.

To reduce the computational complexity of the probability density function in 5-dimensional space, CUDA programming was used to accelerate graphics processing unit (GPU) operations on images. Color information is mapped to a 2D image. The protocol algorithm is used to search for color points with the maximum probability density function value. The color of this point and the corresponding function values are output to the density map and color map, respectively. Fig. 2 depicts the protocol algorithm [20-21].

The constraint algorithm in Fig. 2 follows the CUDA setting for finding the probability density function. For the 3D color space corresponding to the original pixel Block, the maximum value of Thread in 16 color spaces is calculated sequentially, transforming the 3D color space into a 2D color space. Recursively compare the density values of four positions in a two-dimensional color space. The maximum values of the four positions are passed to the next layer of color space until the maximum density value in the three-dimensional color space is finally determined. The density map containing contour information estimated by the spatial color probability density function has unclear boundaries. Therefore, a non-maximum suppression algorithm is adopted to refine the image boundary contour and apply it to IS [22]. Traditional algorithms only use image contour information and ignore image color information, resulting in the inability to specify segmentation levels. The DPC algorithm was applied to IS work to obtain the DPC-IS algorithm. DPC is a clustering algorithm based on center points. The local density ld of the sample points is used to define the clustering center t and the nearest distance md. The local density ld_i of sample point i is illustrated in Eq. (7) [23].

$$ld_i = \sum_j \chi\left(d\left(i,j\right) - d_c\right) \tag{7}$$

In Eq. (7), $d\left(i,j\right)$ stands for the Euclidean distance between samples i and j. d_c stands for the cutoff distance. When $x<0$, $\chi\left(x\right)=1$. Otherwise, the value is 0. The expression of cluster center t_i is illustrated in Eq. (8).

$$t_i = \arg\min_{j:ld_j > ld_i}\left(d\left(i,j\right)\right) \tag{8}$$

In Eq. (8), ld_j refers to the local density of sample point j. The $\arg\min$ function refers to the value of a variable when $d\left(i,j\right)$ reaches the minimum value. The segmentation effect varies with the number of cluster centers, as shown in Fig. 3.

Fig. 3 shows the segmentation results with clustering centers of 2, 5, 7, 12, 14, and 20, respectively. As the number of clustering centers increases, the regions segmented by the DPC-IS algorithm are also increasing. The segmentation details are gradually improving. The nearest distance md_i of sample point i is shown in Eq. (9).

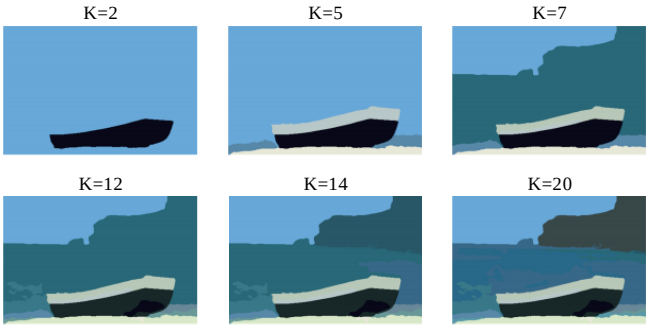


Fig. 3. Schematic diagram of segmentation effect of different clustering centers.

$$md_i = \min_{j:ld_j > ld_i} d\left(i,j\right) \tag{9}$$

When the DPC algorithm is applied to IS, the previously obtained density map containing pixel distribution center information can be directly used as the local density of pixels. The nearest distance of pixel point k is shown in Eq. (10).

$$md_k = \min d_c\left(i,j\right) \tag{10}$$

In Eq. (10), pixels l and k are segmented adjacent. $ld_l > ld_k$. The definition of adjacent segmentation is that if two pixels are adjacent, the segmentation area to which the two pixels belong is adjacent.

B. Construction of FasterR-CNN Algorithm in Image Detection

The CNN model, as a fresh neural network model, is extensively applied in image processing. Unlike traditional neural network models, CNN models add many feature learning components to the existing network. The hidden layer is reasonably divided into convolutional layer and pooling layer, in order to directly use convolutional layer to train image data. The features of interested objects in the graph are extracted, reducing the computational workload of object detection. The CNN model structure is shown in Fig. 4.

Fig. 4 depicts the essential framework of the CNN model. It mainly consists of a series of convolutional layers, incentive layers, pooling layers, and fully connected layers (FCL). The expression of convolutional layer is shown in Eq. (11).

$$x_n^l = f\left(\sum_{i\in G_n} x_m^{l-1} * Kernel_{mn}^l + a^l\right) \tag{11}$$

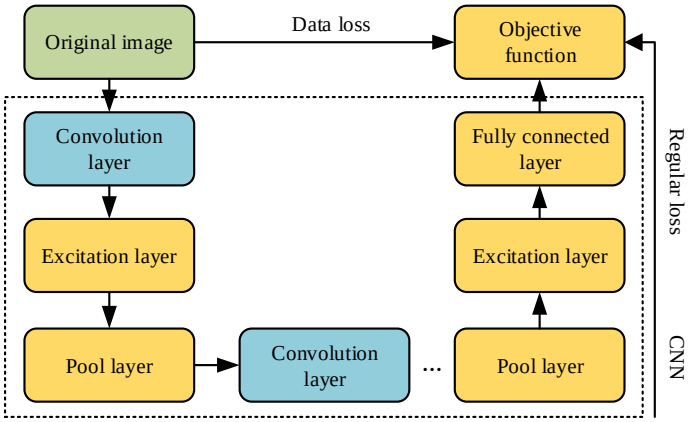


Fig. 4. Schematic diagram of CNN model structure.

In Eq. (11), x_n^l represents the m-th neuron in the l-th layer. *Kernel* refers to the convolutional kernel of the layer. b is the bias term. * represents convolutional operations. G_n represents the number of inputs for the n-th neuron. $f(\square)$ is the nonlinear excitation function. The CNN model has local connectivity. Therefore, each neuron can only perceive a portion of the image, effectively reducing the interaction between neurons and computational complexity. Additionally, the same feature may exist at different locations in the image. The convolutional layer adopts a parameter sharing method, so that each convolutional kernel corresponds to a type of feature, thereby reducing the required parameters for training. The excitation layer introduces nonlinear features into the neural network, enabling it to approximate any nonlinear function. This effectively enhances the feature expression ability of CNN models. The commonly used expression of the excitation function is shown in Eq. (12).

$$\begin{cases} f_S(x) = \dfrac{1}{1+e^{-x}} \\ f_T(x) = \dfrac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \\ f_R(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases} \end{cases} \tag{12}$$

In Eq. (12), $f_S(x) = \dfrac{1}{1+e^{-x}}$ is sigmoid function. $f_T(x) = \dfrac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$ is the Tanh function. $f_R(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases}$ is ReLU function. The output of neurons is mapped to the $(0,1)$ interval through the sigmoid function. However, this function has a soft saturation characteristic, which can cause gradient disappearance in deep networks. The average output value of the Tanh function is closer to 0. It has good convergence, but still cannot solve the gradient disappearance. The ReLU function sets the output values of some neurons to 0, making the neural network sparse. To some extent, the problem of overfitting has been avoided. In addition, the derivative of the positive activation value of the ReLU function is 1, which can effectively avoid gradient disappearance. Therefore, ReLU function is used as the activation function of the excitation layer [24]. The pooling layer commonly uses maximum pooling and mean pooling to reduce the dimensionality of feature maps. The network parameters and computational complexity are reduced to facilitate feature compression, as shown in Fig. 5.

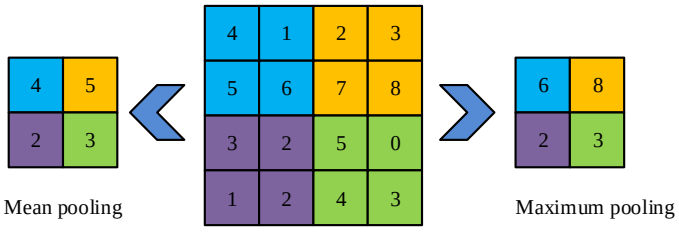


Fig. 5. Mean pool and maximum pool.

The feature data obtained from mean pooling in Fig. 5 is relatively sensitive to texture information. Maximizing pooling places more emphasis on background information. The function of the FCL is to integrate the features of the input image. The feature maps generated by convolutional layers are mapped into fixed length feature vectors to complete image classification tasks. The calculation of the FCL is shown in Eq. (13).

$$x_n^l = f\left(\sum_{i \in G_n} x_m^{l-1} * w_{m,n} + a_n\right) \tag{13}$$

In Eq. (13), x_n^l refers to the output neuron of the FCL. $w_{m,n}$ and a_n represent the weights and offsets between neurons, respectively. The image classification task requires adding a classifier after the high-dimensional feature vectors of the FCL. The feature vectors of the FCL are mapped to the probability of each category. The category with the highest probability is the classification result. The expression of function $g_\theta(x)$ is shown in Eq. (14).

$$g_\theta(x^i) = \frac{1}{\sum_{n=1}^{k} e^{\theta_n^T x^n}} \begin{bmatrix} e^{\theta_1^T x^n} \\ e^{\theta_2^T x^n} \\ \cdots \\ e^{\theta_n^T x^n} \end{bmatrix} \tag{14}$$

In Eq. (14), $\theta = \left[\theta_1^T \theta_2^T \cdots \theta_n^T\right]$ is the learning parameter of the classifier, including the Softmax cost function $F(\theta)$ of the weight penalty term, as shown in Eq. (15).

$$F(\theta) = -\frac{1}{i}\left[\sum_{m=1}^{k}\sum_{n=1}^{k}\{y^m = n\}\log\frac{e^{\theta_n^T x^m}}{\sum_{l=1}^{k} e^{\theta_l^T x^m}}\right] + \frac{\beta}{2}\sum_{m=1}^{k}\sum_{n=0}^{j}\theta_{mn}^2 \tag{15}$$

In Eq. (15), i is the samples. β is the coefficient of the regularization term. In the training process, CNN uses the Gradient descent to minimize $F(\theta)$, so as to find the optimal solution of parameter weight θ. Traditional regional CNN models consume a long amount of computation time when performing object detection. It is easy to cause computational redundancy and a large amount of memory consumption, resulting in slow target detection speed. To improve the detection accuracy and speed of graphics, the FasterR-CNN has been introduced. The FasterR-CNN algorithm is a deep learning algorithm that introduces regional recommendation networks to improve fast regional CNN networks. This algorithm utilizes convolutional network training to achieve target detection without any other algorithm operations. The selective search algorithm in regional recommendation networks is different from that in traditional regional CNN models. This method uses the feature map extracted from the last layer of CNN as the input feature. Convolutional kernels are used to move within the feature map and output suggestion boxes. Every time the convolution kernel moves, it will take the central position of the window as the base point to build nine anchor windows with different sizes and different Aspect ratio. It is mapped into the image to form a feature vector. Then the FCL is applied for classification and positioning. The FasterR-CNN algorithm structure is shown in Fig. 6.

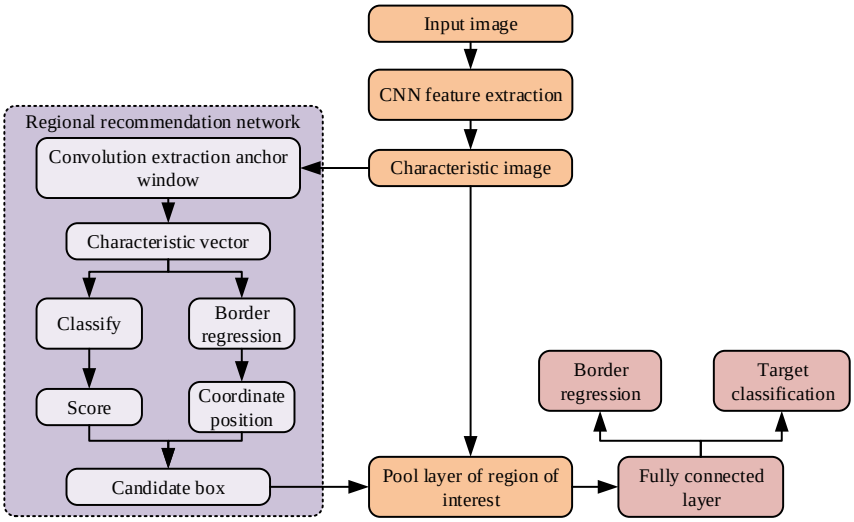


Fig. 6. FasterR-CNN algorithm flow diagram.

The algorithm flow in Fig. 6 is as follows. Firstly, the input image is extracted using CNN for feature extraction. Then input the extracted feature images into the region recommendation network for candidate border generation. The region border and feature map classify and locate the objective function through the pooling layer of interest in the fast region CNN network. The final image target detection effect is obtained.

IV. Result Analysis of 3DPC-IS Algorithm and FasterR-CNN Algorithm

To better validate the feasibility of the DPC-IS algorithm and FasterR-CNN algorithm, multiple control groups are set up in the BSDS500 database and Matlab simulation software for comparative experiments. This chapter focuses on the performance of DPC-IS algorithm and FasterR-CNN algorithm.

A. Performance Analysis of DPC-IS Algorithm

To prove the effectiveness of the DPC-IS algorithm, comparative experiments are designed using Simple Linear Iterative Clustering (SLIC), Fast Partitioning of Vector Value Images (FPVVI) algorithm, Normalized Cuts (NC) algorithm, and DPC-IS algorithm. The performance indicator is IS accuracy. The experimental platform is the BSDS500 database.

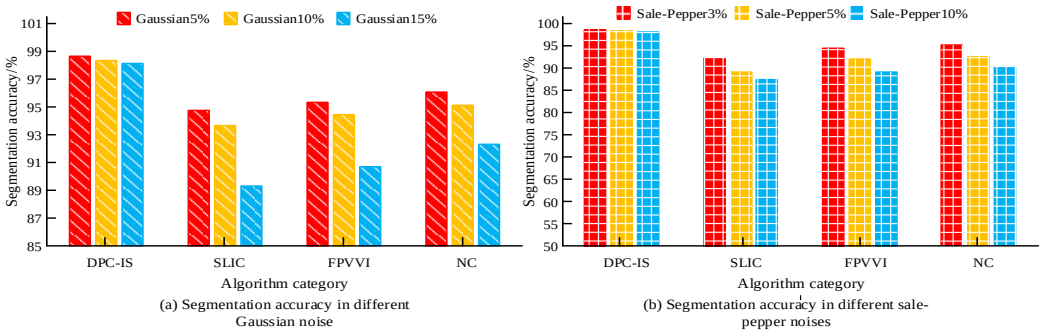


Fig. 7. Segmentation accuracy of four algorithms under different noise categories.

Fig. 7 depicts the segmentation accuracy of four algorithms in different noise categories. Fig. 7(a) depicts the segmentation accurac of four algorithms in different Gaussian noise categories. When the Gaussian noise is 5%, the segmentation accuracy of PDC algorithm is 98.65%. The SLIC, FPVVI, and NC algorithms are 94.75%, 95.33%, and 96.07%, respectively. When the Gaussian noise is 15%, the PDC, SLIC, FPVVI and NC algorithms are 98.13%, 89.31%, 90.69% and 92.31% respectively. Fig. 7(b) illustrates the segmentation accuracy of four algorithms in different Salt-and-pepper noise categories. When Salt-and-pepper noise is 3%, the segmentation accuracy of PDC algorithm is 98.45%. The SLIC, FPVVI, and NC algorithms are 92.15%, 94.32%, and 94.99%, respectively. When the Salt-and-pepper noise is 10%, the PDC, SLIC, FPVVI, and NC algorithms are 97.89%, 87.32%, 89.05%, and 90.12%, respectively. The research results show that as the media of Visual communication, the PDC algorithm has superior performance in the processing.

Fig. 8. Segmentation effect of four algorithms on some images.

The segmentation effects of the four algorithms on some images are shown in Fig. 8, Fig. 8(a) to Fig. 8(e) show the segmentation effects of the original image, DPC-IS algorithm, FPVVI algorithm, SLIC algorithm, and NC algorithm, respectively. The DPC-IS algorithm can effectively segment images. Compared to FPVVI, SLIC, and NC algorithms, the DPC-IS algorithm filters the background while retaining the main salient regions. Important areas are displayed separately from the background, making the visual communication effect more prominent. To assess the effectiveness in segmenting images, three performance indicators were introduced: Variance of Information (VI), Probabilistic Rand Index (PRI), and Segmentation Coverage (SC). VI is the difference between the segmentation and the benchmark result. The smaller the value, the better. The higher the values of PRI and SC, the better the performance.

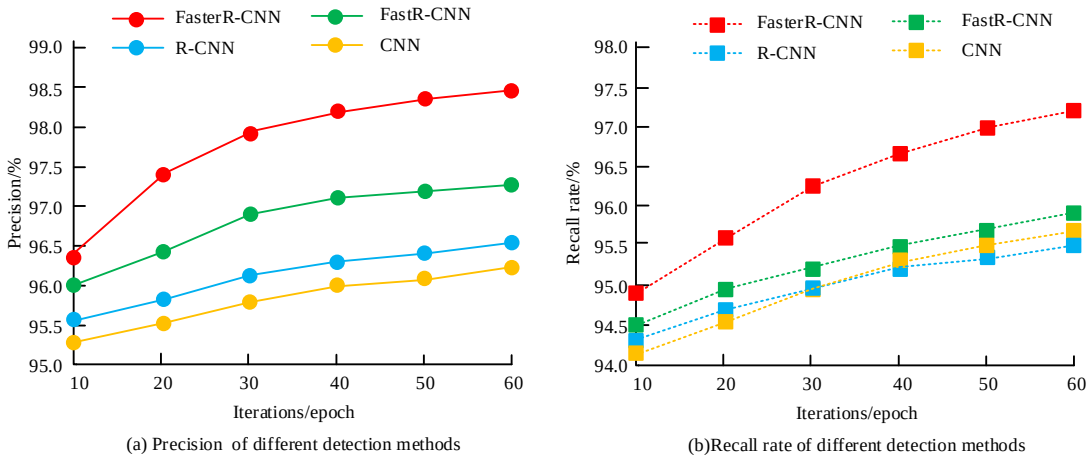
Table I shows the PRI, VI, and SC results of the two algorithms. The Optimal Dataset Scale (ODS) and Optimal Image Scale (OIS) represent the optimal result thresholds within the database and the optimal result thresholds for each image, respectively. The PRI values of DPC-IS algorithm under ODS and OIS are 0.831 and 0.835, respectively, which are higher than those of SLIC algorithm's 0.743 and 0.749. The VI value of the DPC-IS algorithm under ODS is 1.892. The VI value of the SLIC algorithm is 2.257. The VI value of the PDC algorithm under OIS is 2.074. The VI value of the SLIC algorithm is 2.420. The SC values of DPC-IS and SLIC algorithm under ODS are 0.564 and 0.508, respectively. The SC values under OIS are 0.588 and 0.514. From the table, DPC-IS algorithm has good segmentation performance in image processing and performs well in the field of Visual communication.

TABLE I. PRI, VI AND SC RESULTS OF TWO ALGORITHMS

Evaluating indicator	Optimal threshold	DPC-IS	SLIC
PRI	ODS	0.831	0.743
	OIS	0.835	0.749
VI	ODS	1.892	2.257
	OIS	2.074	2.420
SC	ODS	0.564	0.508
	OIS	0.588	0.514

B. *Performance Analysis of FasterR-CNN Algorithm in Image Detection*

To verify the availability, of FastR-CNN in image detection, FastR-CNN algorithm, R-CNN algorithm, CNN algorithm, and FastR-CNN algorithm were used for comparative experiments. The performance indicators are accuracy, recall, and F-measure values. The experimental platform is Matlab simulation software.



(a) Precision of different detection methods (b)Recall rate of different detection methods

Fig. 9. Precision and recall rate results of four algorithms.

Fig. 9 illustrates the accuracy and recall results of the four algorithms. As the iterations increase, the accuracy and recall of the four algorithms all increase. Fig. 9(a) shows the accuracy variation curves of the four algorithms. When the iteration is 10, the accuracy of the FasterR-CNN is 96.41%. The accuracy rates of FastR-CNN, R-CNN, and CNN algorithms are 96.03%, 95.54%, and 95.32%, respectively. When the iteration is 60,the accuracy of FastR-CNN is 98.49%, which is higher than the 96.87%, 96.35%, and 96.04% of FastR-CNN, R-CNN, and CNN algorithms. Fig. 9(b) shows the recall rate variation curves of the four algorithms. When the number of iterations is 10, the recall rate of FasterR-CNN algorithm is 94.95%. The recall rates of FastR-CNN, R-CNN, and CNN algorithms are 94.48%, 94.31%, and 94.11%, respectively. When the iteration is 60, the recall rate of FastR-CNN algorithm is 97.29%, which is higher than the 95.60%, 95.38%, and 95.24% of FastR-CNN, R-CNN, and CNN algorithms. From the figure, FasterR-CNN has high detection accuracy, which is robust in the field of Visual communication.

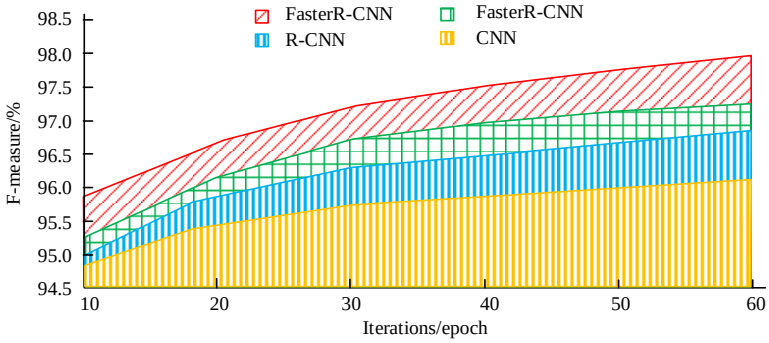
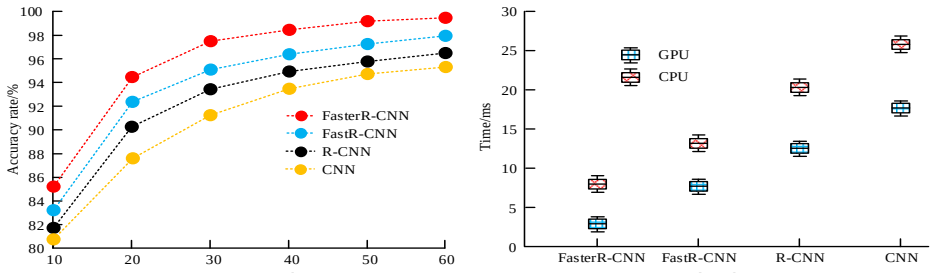


Fig. 10. F-measure results of four algorithms.

The F-measure values of the four algorithms are illustrated in Fig. 10. The F-measure increases with the iteration. When the iteration is 10, the F-measure value of the FasterR-CNN algorithm is 95.90%. The F-measure values of FastR-CNN, R-CNN, and CNN algorithms are 95.26%, 94.99%, and 94.74%, respectively. When the iteration is 60, the F-measure value of the FasterR-CNN algorithm is 97.77%. The F-measure of the FastR-CNN algorithm is 96.87%. The F-measure of the R-CNN algorithm is 96.49%. The F-measure value of the CNN algorithm is 95.51%. From the figure, the FasterR-CNN algorithm has robust performance and good performance.



(a) Accuracy results of different algorithms (b) Time consuming of different algorithms in GPU environment and CPU environment.

Fig. 11. Performance results of four algorithms.

Fig. 11 illustrates the experimental results of four algorithms. Fig. 11(a) shows the accuracy results of the four algorithms. As iterations increase, the accuracy of all four algorithms is improving. When the iteration is 60, the accuracy of FastR-CNN is 98.18%, FastR-CNN algorithm is 96.52%, R-CNN algorithm is 95.66%, and CNN algorithm is 93.33%. Fig. 11(b) shows the average time taken by four algorithms to detect a single image in GPU and CPU environments, respectively. The FasterR-CNN algorithm in GPU environment takes 2.94ms. The FastR-CNN algorithm takes 6.28ms. The R-CNN algorithm takes 12.04ms. The CNN algorithm takes 16.21ms. The FasterR-CNN algorithm takes 8.71ms in CPU environment. The FastR-CNN algorithm takes 13.05ms. The R-R-CNN algorithm takes 19.87ms. The R-CNN algorithm takes 25.19ms. The FasterR-CNN algorithm has higher accuracy. The real-time performance is higher than the other three algorithms in both GPU and CPU environments.

V. Conclusion

As a common carrier of Visual communication, image processing with artificial intelligence algorithm has gradually become a development direction. Common image processing technologies include IS technology and image detection technology. Traditional IS technology often only involves the contour information of the image, while ignoring color information, resulting in unclear segmentation levels. Traditional image detection techniques have long computational time consumption, redundant computation, and poor detection efficiency. To address the above issues, the DPC based IS algorithm and FasterR-CNN algorithm are used for image segmentation and detection, respectively. According to the findings, the segmentation accuracy of SLIC is 89.31% and 87.32% respectively in the case of 15% Gaussian noise and 10% Salt-and-pepper noise. The DPC-IS algorithm is 98.13% and 97.89% respectively, significantly higher than SLIC. In the ODS, the PRI and SC values of the PDC algorithm are 0.831 and 0.564, respectively, which are 0.088 and 0.056 higher than the SLIC. The VI value of the PDC is 1.892, which is 0.365 less than the SLIC algorithm. When the iteration is 60, the accuracy, recall, and F-measure values of FastR-CNN algorithm are 98.49%, 97.29%, and 97.77%, respectively, which are 1.62%, 1.69%, and 0.90% higher than FastR-CNN algorithm. In summary, the DPC-IS algorithm and FasterR-CNN algorithm proposed in the study have superior performance, effectively improving segmentation and detection accuracy. However, there are still shortcomings in the research. Firstly, the process of mapping a five dimensional space to a two-dimensional space using maximum mapping may result in data loss, which in turn affects the stability of the algorithm. Secondly, the performance of the proposed method under extreme noise conditions or highly complex image environments still needs further validation. In addition, further efforts are needed to optimize the stability of these algorithms for practical applications. Future research should focus on integrating more complex spatial transformation techniques, such as top-k mapping, to reduce data loss during dimensionality reduction. In addition, exploring the application of these algorithms in dynamic and real-time environments, such as video processing, will provide valuable insights into their scalability. Meanwhile, studying their effectiveness in a wider range of image types, such as medical or satellite images, may also open up new avenues for enhancing visual communication technology.

References

[1] Zomay Z, Keskin B, Ahin C. Grsel letiim Tasarm Blümü rencilerinin Sektrel Logolardaki Renk Tercihleri - Color Preferences of Visual Communication Design Students in Sectoral Logos. OPUS Uluslararası Toplum Araştırmaları Dergisi, 2021, 17(37):4181-4198.

[2] Hidayat I, Ali M Z, Arshad A. Machine Learning-Based Intrusion Detection System: An Experimental Comparison. Journal of Computational and Cognitive Engineering, 2022, 2(2):88-97.

[3] Liu X. Application of cloud-based visual communication design in Internet of Things image. Soft computing: A fusion of foundations, methodologies and applications, 2020, 24(11):8041-8050.

[4] Rai M, Goyal S, Pawar M. An enhanced digital image watermarking technique using DWT-HD-SVD and deep convolutional neural network[J]. International Journal of Critical Computer-Based Systems, 2023, 10(4): 269-286.

[5] Zhang J, Zhou Y, Xia K, Jiang Y, Liu Y. A novel automatic image segmentation method for Chinese literati paintings using multi-view fuzzy clustering technology. Multimedia Systems, 2020, 26(1):37-51.

[6] Amna M, Imen W, Sayadi F E, Atri M. Fast intra-coding unit partition decision in H.266/FVC based on deep learning. Journal of Real-Time Image Processing, 2020, 17(6):1971-1981.

[7] Vani R, Rajan K S. Effective satellite image enhancement based on the discrete wavelet transform. International Journal of Business Information Systems, 2020, 33(4):446-471.

[8] Zhao F, Lu H, Zhao W, Yao L. Image-Scale-Symmetric Cooperative Network for Defocus Blur Detection. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(5):2719-2731.

[9] Schonsheck S C, Dong B, Lai R. Parallel transport convolution: Deformable convolutional networks on manifold-structured data[J]. SIAM Journal on Imaging Sciences, 2022, 15(1): 367-386.

[10] Shi Q, Yin S, Wang K, Teng L, Li H. Multichannel convolutional neural network-based fuzzy active contour model for medical image segmentation. Evolving Systems, 2022, 13(4):535-549.

[11] Liu X, Zhang Y, Jing H, Wang L, Zhao S. Ore image segmentation method using U-Net and Res_Unet convolutional networks. RSC Advances, 2020, 10(16):9396-9406.

[12] Ji Y, Jiang X. Active contour model for image segmentation based on salient fitting energy. International Journal of Information and Communication Technology, 2021, 19(2):219-230.

[13] Li Y, Cao G, Wang T, Cui Q, Wang B. A novel local region-based active contour model for image segmentation using Bayes theorem. Information Sciences, 2020, 506:443-456.

[14] Biswas S, Hazra R. Active contours driven by modified LoG energy term and optimised penalty term for image segmentation. IET Image Processing, 2020, 14(13):3232-3242.

[15] Chen J. Automatic Vehicle License Plate Detection using K-Means Clustering Algorithm and CNN. Journal of Electrical Engineering and Automation, 2021, 3(1):15-23.

[16] Vinolin V, Sucharitha M. Taylor-rider-based deep convolutional neural network for image forgery detection in 3D lighting environment. Data technologies and applications, 2022, 56(1):103-131.

[17] Luo J Q, Fang H S, Shao F M, Zhong Y, Hua X. Multi-scale traffic vehicle detection based on faster R–CNN with NAS optimization and feature enrichment. Defence Technology, 2021, 17(4):1542-1554.

[18] Ramaraj P. A Neural Network in Convolution with Constant Error Carousel Based Long Short-Term Memory for Better Face Recognition. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 2021, 12(2):2042-2052.

[19] Joy F, Vijayakumar V. MULTIPLE OBJECT DETECTION IN SURVEILLANCE VIDEO WITH DOMAIN ADAPTIVE INCREMENTAL FAST RCNN ALGORITHM. Indian Journal of Computer Science and Engineering, 2021, 12(4):1018-1026.

[20] Venetis I E, Vasso S, Stathis S, Efstratios G. Multivariable inversion using exhaustive grid search and high-performance GPU processing: a new

perspective. Geophysical Journal International, 2020, 221(2):905-927.

[21] Arieth R M, Anuradha K, Harika B. Congestion Management of CGSTEB Routing protocol Using K-means Algorithm in Wireless Sensor Network. ECS transactions, 2022, 107(1):13147-13154.

[22] Harakannanavar S S, Sameer S R, Kumar V, Behera SK, Amberkar AV, Puranikmath VI. Robust video summarization algorithm using supervised machine learning. Global Transitions Proceedings, 2022, 3(1):131-135.

[23] Shi Y, Shen H. Anomaly Detection for Network Flow Using Immune Network and Density Peak. International Journal of Network Security, 2020, 22(2):337-346.

[24] Chin W L, Zhang Q, Jiang T. Low-complexity neuron for fixed-point artificial neural networks with ReLU activation function in energy-constrained wireless applications. IET communications, 2021, 15(7):917-923.

Revolutionizing Rice Leaf Disease Detection: Next-Generation SMOREF-SVM Integrating Spider Monkey Optimization and Advanced Machine Learning Techniques

Avip Kurniawan[1], Tri Retnaningsih Soeprobowati[2], Budi Warsito[3]
Department of Information System-School of Postgraduate Studies, Universitas Diponegoro, Semarang, Indonesia[1]
Department of Informatics-Faculty of Engineering, Universitas Krisnadwipayana, Jakarta, Indonesia[1]
Department of Biology-Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia[2]
Cluster for Paleolimnology, Universitas Diponegoro, Semarang, Indonesia[2]
Department of Statistics-Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia[3]

***Abstract*—Leaf diseases pose a significant challenge to rice productivity, which is critical as rice is a staple food for over half of the world's population and a major agricultural commodity. These diseases can lead to severe economic losses and jeopardize food security, particularly in regions heavily reliant on rice farming. Traditional detection methods, such as visual inspection and microscopy, are often inadequate for early disease identification, which is crucial for effective management and minimizing yield loss. This presentation introduces SMOREF-SVM, a novel approach that combines Spider Monkey Optimization (SMO) with Random Forest (RF) and Support Vector Machine (SVM) to improve the classification of rice leaf diseases. The innovation of SMOREF-SVM lies in its use of SMO for effective feature optimization, which selects the most relevant features from complex disease patterns, and its dual-classification framework using RF and SVM. Results demonstrate that SMOREF-SVM achieves an average accuracy of 98%, significantly outperforming standard SVM methods, which achieve around 90%. SMOREF-SVM also improves key metrics, including Precision, Recall, and F1 Score, by 5-10% for diseases with fewer samples, reaching Precision of 94%, Recall of 92%, and F1 Score of 93%. Additionally, ROC curve analysis shows an enhanced Area Under the Curve (AUC), approaching 0.98 for more disease classes, compared to 0.85 with traditional methods. This makes SMOREF-SVM a valuable tool for early and accurate disease detection, offering the potential to improve crop productivity and sustainability, addressing the critical challenges of disease management in agriculture.**

Keywords*—*SMOREF-SVM; rice leaf disease; classification; Spider Monkey Optimization (SMO); machine learning; image processing

I. INTRODUCTION

Leaf diseases are one of the major problems in plant growth that can cause significant obstacles in the agribusiness sector, as well as negatively impacting a country's agricultural production. The main causes of leaf diseases involve various types of pathogens such as bacteria, fungi, viruses [1], and other natural infectious organisms, which can attack plants at various stages of their life cycle. To detect and classify the types of leaf stress, there are several approaches that can be used. The first approach is to observe directly with the naked eye, which although simple, is often ineffective and inefficient. This visual observation tends to be slow and less accurate, especially in detecting early signs of infection on leaves. Another alternative is to use special instruments, such as microscopes, to observe disease symptoms. However, this method is also not ideal because it takes a long time, so that preventive measures cannot be taken quickly before the disease spreads further.

Therefore, many studies are now turning to a more sophisticated approach, namely the application of machine learning (ML) algorithms to detect and classify leaf diseases. The application of these ML techniques provides a faster and more accurate solution in identifying leaf diseases at an early stage, when symptoms of infection may not be clearly visible to the human eye. By leveraging the capabilities of image processing (IP) and computer vision, these techniques enable more effective automated identification of disease symptoms on leaves [2]. Overall, leaf disease identification using ML and image processing technologies is one of the most promising and challenging areas of research. This is because early detection of disease is a crucial step in preventing the spread of infection and greater losses to agriculture. With the advancement of this technology, it is expected that the agricultural sector can utilize new innovations to increase productivity and reduce the negative impact of leaf diseases on crops [1] [2]. Farmers' knowledge of identifying diseases in rice plant leaves is generally derived from traditional knowledge passed down through generations [3]. Beginner rice farmers often face difficulties in recognizing these diseases due to their limited understanding of the common afflictions that affect rice plants, especially their leaves. These diseases can cause significant losses, such as reduced crop yields or even total crop failure when the disease becomes severe and difficult to control [4]. Common diseases that frequently affect rice plants include Bacterial Leaf Blight, Tungro, Leaf Blast, Sheath Blight, False Smut, and Grassy Stunt [1]. Early identification and prevention of these diseases are crucial to mitigate losses, such as reduced productivity due to disease

outbreaks. Currently, image processing and machine learning technologies are increasingly utilized for rapid and efficient detection of diseases in rice plant leaves [5].

The general steps in this process involve several stages, including data acquisition, preprocessing, segmentation, feature extraction, and classification [6]. Numerous studies have been conducted to develop methods for identifying diseases in rice plant leaves. One approach involves identifying diseases by applying feature extraction techniques such as discrete wavelet transform, scale-invariant feature transform, and gray-level co-occurrence matrix, followed by classification using algorithms like K-nearest neighbor, backpropagation neural network, Naive Bayesian, and multiclass SVM. The results of the research indicated that the multiclass SVM algorithm provided the highest accuracy performance at 98.63% [7]. Another study focusing on classifying leaf diseases like Paddy Blast, Narrow Brown Spot, Brown Spot, and healthy leaves involves multiple steps, including color feature extraction to obtain binary image values for area (A) and perimeter (P), shape feature extraction to acquire values for rectangularity, compactness, elongation, and roundness, and texture feature extraction using GLCM to measure contrast, uniformity, entropy, inverse difference, and linear correlation. The SVM algorithm is then used for classification, and the study demonstrated an average accuracy performance of 92.5% [8]. Further research on detecting healthy leaves, Brown Spot, and Narrow Brown Spot utilized feature extraction methods such as RGB color slicing values, edge feature values, and color histogram values, resulting in an average accuracy performance of 89% [9]. Additionally, another study classified diseases like Bacterial Leaf Blight, Leaf Blast, and Brown Spot on rice plants using color feature extraction with intensity moment values (Mean, Median, Mode, Midrange, Range, IQR, and Standard Deviation). Classification was performed by automatically selecting the best algorithm among Random Forest, KNN, and SVM, achieving the highest accuracy performance of 91.47% [10]. Numerous articles have been published on research studies in leaf disease detection, which have been adopted by universities, private organizations (Pvt. Org.), government organizations (Govt. Org.), industry, and collaborative efforts between universities and private entities (U&P). These studies highlight a growing interest across different sectors in applying advanced machine learning (ML) and deep learning (DL) models to enhance agricultural productivity and efficiency.

Many universities have selected both fundamental and cutting-edge ML and DL models to address their specific research objectives in leaf disease detection [10][11]. This includes adopting models for early detection, classification, and analysis of plant diseases to support agricultural education and research. Similarly, government organizations, private entities, and industries have funded numerous research projects, recognizing the critical need to improve disease management practices in agriculture. The main motivation behind these efforts is to develop technologies that provide farmers with efficient, reliable, and user-friendly tools for detecting and managing leaf diseases, thereby minimizing crop loss and enhancing overall agricultural output. Industry 4.0, the fourth industrial revolution, is crucial in this context. Integrating advanced technologies such as the Internet of Things (IoT), artificial intelligence (AI), big data analytics, robotics, and cloud computing into various processes, including agriculture. In the realm of leaf disease detection, Industry 4.0 provides a framework for using these technologies to monitor plant health in real-time, automate the detection process, and optimize resource use [12] [13]. For example, IoT devices equipped with sensors can collect data on environmental conditions and plant health, while AI and ML algorithms analyze this data to identify disease patterns and predict outbreaks. Digital transformation, on the other hand, represents a broader shift towards using digital technologies to fundamentally reshape how businesses operate and deliver value to customers. In agriculture, digital transformation leverages tools like cloud computing, AI, machine learning, and big data analytics to enhance the efficiency and accuracy of leaf disease detection. This transformation involves the adoption of digital platforms for data sharing, real-time monitoring systems for early disease detection, and decision-support systems to guide farmers on the best course of action based on data-driven insights [14].

In the context of leaf disease detection, Industry 4.0 and digital transformation are key drivers behind the adoption of advanced technologies. Their primary focus is on fostering innovation, improving operational efficiency, and providing farmers with sophisticated yet accessible tools to address plant diseases [15]. These efforts are crucial for advancing sustainable agricultural practices, reducing crop losses, and ultimately contributing to global food security. Image processing and machine learning technologies are continually evolving as effective tools for detecting and classifying diseases in rice plant leaves, offering new hope for enhancing agricultural resilience and productivity. This study focuses on rice plants, a staple crop that plays a vital role in food security. The harvested paddy must undergo a milling process to produce rice ready for consumption. Despite a steady increase in population, rice production has been declining, partly due to diseases affecting rice plants [16]. Therefore, preventive measures against rice plant diseases are essential to mitigate productivity losses, starting with the identification of various diseases that commonly affect rice plants. Some prevalent diseases include Leaf Blast, Tungro, Sheath Blight, Grassy Stunt, Bacterial Leaf Blight, and False Smut. Understanding these diseases enables the implementation of appropriate preventive actions [17]. Disease identification can be performed through visual inspection or laboratory analysis by closely examining the symptoms present on the leaf surfaces. The Table I provides an overview of the symptoms or signs commonly observed in rice plant leaves [18].

The research on rice leaf disease classification has explored various advanced methods and algorithms to achieve high accuracy. For example, compared different Convolutional Neural Network (CNN) architectures, including VGG16, MobileNet, Xception, and ResNet34, using transfer learning techniques [19]. Their study found that the ResNet34 architecture achieved the highest accuracy, reaching 97.50%. Similarly, proposed a GCL model that combines Generative Adversarial Networks (GAN), CNN, and Long Short-Term Memory (LSTM). In their approach, GAN was used for data

augmentation, CNN for feature extraction to differentiate disease information on rice leaves, and LSTM for classification, resulting in an average accuracy of 97% [20] (Fig. 1).

TABLE I. COMMON DISEASES IN RICE PLANTS AND THEIR SYMPTOMS

Type of Disease		Symptoms
a	Symptoms of Leaf Blast	Spots with pointed tips; the center of the spots is grayish-white with a halo area around it.
b	Symptoms of Tungro Disease	Symptoms begin with leaf discoloration and stunted growth.
c	Symptoms of Bacterial Leaf Blight	Brown spots characterized by small, dark brown spots; spots that develop with brown edges and pale yellow, dirty white, brown, or gray centers.
d	Symptoms of Grassy Stunt	Initial symptoms include small, dark brown spots; advanced symptoms show larger spots with dirty white centers.
e	Symptoms of Sheath Blight	Irregularly shaped spots; the edges of the spots are reddish, and the center of the spots is light brown.
f	Symptoms of False Smut	The disease-causing organism develops inside the rice husk, transforming the endosperm into a large fungal sclerotium that protrudes outside, appearing golden yellow.

(a) (b) (c) (d) (e) (f)

Fig. 1. Typical diseases encountered in rice plants [21].

Deep Convolutional Neural Network (DCNN) combined with a Cuckoo Search (CS) algorithm, called DCNN-CS, to classify rice leaf diseases. The CS algorithm helps minimize errors in the classification process, achieving the highest accuracy of 99% [22], and then employed K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) algorithms to classify rice leaf diseases, with KNN yielding an average accuracy of 95% and SVM achieving 98% [23]. SMOREF-SVM (Spider Monkey Optimization with Random Forest and Support Vector Machine) is proposed as a novel approach to improve the accuracy of rice leaf disease classification. This approach has two main motivations: 1) There are still challenges in effectively recognizing rice leaf diseases, which is the main focus of this study. SMOREF-SVM combines feature optimization and dual classification techniques to address this challenge. 2) Based on the complexity of rice leaf disease spot patterns, the Spider Monkey Optimization (SMO) algorithm is applied to select the most relevant features, while Random Forest (RF) and Support Vector Machine (SVM) are used for a more accurate classification stage. This paper makes two main contributions: 1) By combining SMO, RF, and SVM, the proposed method can capture complex patterns of rice leaf disease data, improving detection accuracy and efficiency. The dual classification framework used not only enables the correct identification of diseases but also significantly reduces computational costs. 2) Experimental results show that the combination of SMO with RF and SVM not only accelerates the convergence of the model training process but also produces higher accuracy compared to classical classification methods, such as using Random Forest or SVM separately. Thus, the SMOREF-SVM approach introduces an innovative step in rice leaf disease management, ensuring early detection and more effective management in the field.

II. RESEARCH METHOD

This study involves several key phases to achieve the optimal final result. The first phase is data preprocessing, where raw data is cleaned and processed to ensure quality and consistency before further analysis. The second phase focuses on data extraction, aiming to capture essential features or information from the preprocessed dataset. In the third phase, process optimization is carried out using the Sequential Minimal Optimization (SMO) method, which is then fine-tuned with an 80:20 split between training and test data. After the optimization process, the resulting data is classified using the Random Forest and Support Vector Machine (SVM) algorithms to obtain more accurate predictions. The fourth phase compares the results of the optimized classifications with those from the previous processes to determine which method yields the most precise outcomes. This approach allows researchers to evaluate and compare the effectiveness of different methods and identify the best strategy to maximize data classification accuracy. A more detailed explanation of each step can be found in Fig. 2.

A. Data Preprocessing

The data acquisition process began by collecting 6,000 images of rice leaves, consisting of 1,000 images of healthy leaves and 5,000 images of leaves with various types of diseases. The types of diseases documented include Bacterial Leaf Blight, Blast, Tungro, Sheath Blight, Grassy Stunting, and False Scorch, each with 1,000 images. After the data was collected, the next step was to label each image according to the type of disease. These labels were categorized into seven classes, namely six classes representing various rice leaf diseases and one class for healthy leaves. This labeling is important because it used as classes in the process of creating a classification model, which aims to accurately identify and distinguish between healthy leaves and infected leaves. Thus, this well-labeled dataset structure supports model learning in recognizing the specific characteristics of each type of disease, thereby increasing the accuracy and effectiveness in the process of detecting and classifying rice leaf diseases. After completing the data acquisition stage and collecting a dataset

consisting of images of healthy rice leaves and images of leaves infected with various types of diseases, the next step is to preprocess the dataset.

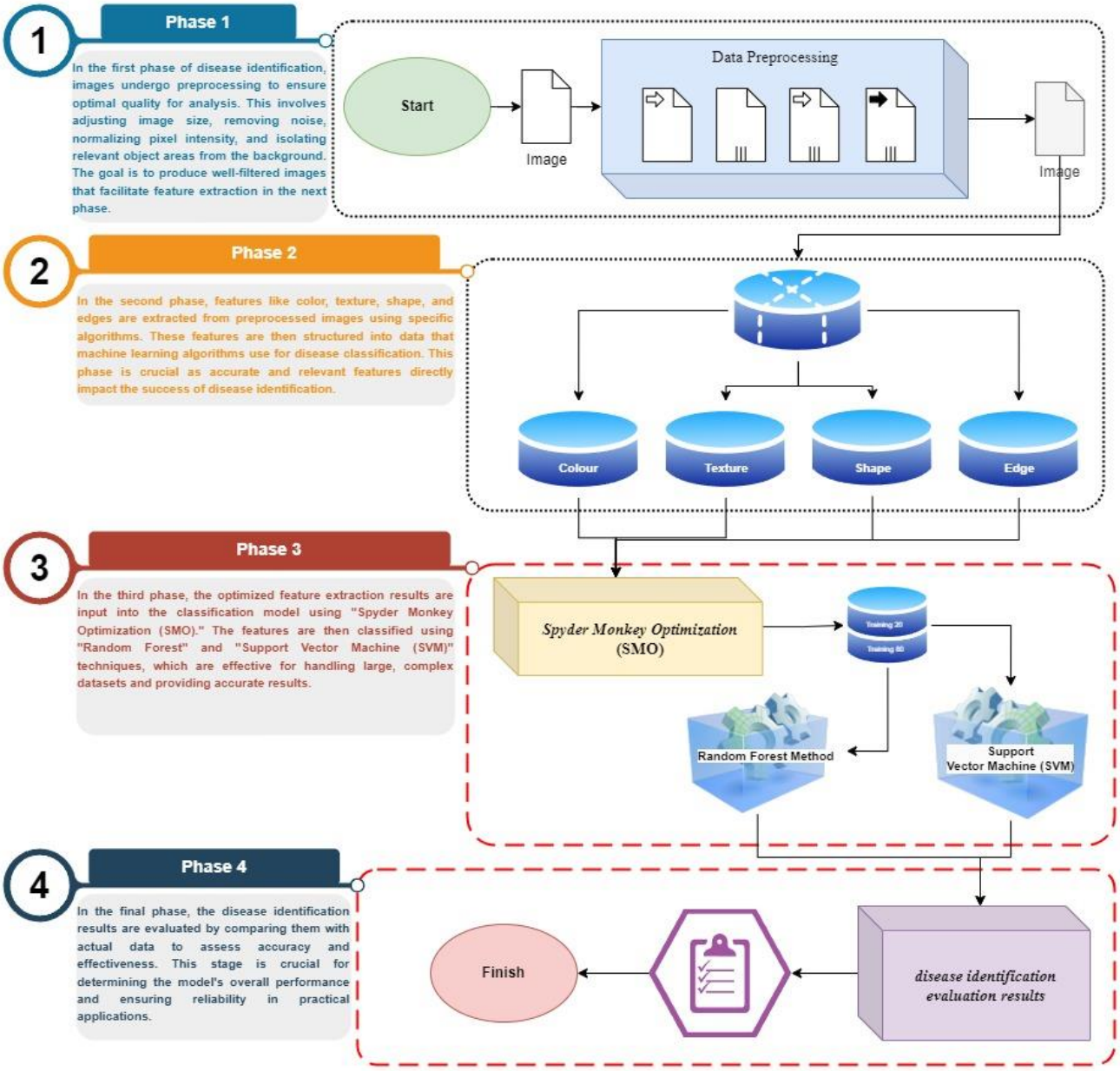


Fig. 2. Proposed method.

This preprocessing involves two main stages, namely resize and cropping. In the first stage, the original images are resized to 256x256 pixels to ensure that the entire dataset has a uniform size, thus facilitating the analysis and classification process. Then, cropping is performed to remove irrelevant or unwanted parts of the image, so that the focus can be directed to more important areas in the image. This step helps highlight the key features needed to improve the accuracy of the model in recognizing and classifying various types of diseases in rice leaves.

B. Data Extraction

The result of the pre-processing stage is an image dataset that has been standardized in size and cleaned from noise interference. After that this image dataset enters the feature extraction stage, which aims to obtain values that reflect the specific characteristics of the image, such as color, texture, edge, and other features. The feature extraction process usually involves several main aspects, such as the extraction of color features, texture, shape, and edge characteristics [21]. In the context of color feature extraction, the approaches used include calculating the average color value, color standard deviation, and color skewness [21]. This feature extraction is an important technique for obtaining representative values from an image dataset [24]. This calculation can be done using several equations, such as the following: for the average color value, the equation can be used:

$$m = \frac{1}{M \times N} \sum_{x=1}^{M} \sum_{y=1}^{N} M_{xy} \quad (1)$$

To calculate the standard deviation of color, the equation is used:

$$SD = \sqrt{\frac{1}{M \times N} \sum_{x=1}^{M} \sum_{y=1}^{N} (M_{xy} - m)^2} \quad (2)$$

The kurtosis value can be calculated by:

$$SK = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} (M_{xy} - m)^3}{(M \times N) \times SD^3} \quad (3)$$

The feature extraction process is not only limited to color but also includes the extraction of shape features, such as solidity, eccentricity, diameter, area, center of mass, minor axis length, and major axis length [24], [25]. To obtain values from the extraction of shape features such as area, aspect ratio, orientation, perimeter, and major and minor axis lengths, the following equation is used [26]:

$$Area = \sum_{x=1}^{M} \sum_{j=1}^{N} A[i, j] \quad (4)$$

$$Aspect\ ratio = \frac{width}{height} \quad (5)$$

$$Orientation = tan^{-1}\left(\frac{y}{x}\right) \quad (6)$$

$$Major\ axis\ and\ minor\ axis = x_1 + x_2\sqrt{(x_1 + x_2)^2 - d} \quad (7)$$

By using this approach, the feature extraction process can effectively identify and analyze various important characteristics in an image dataset, which can then be used for more accurate classification and pattern recognition.

C. Spyder Monkey Optimization (SMO)

Spider Monkey Optimization (SMO) is a relatively new algorithm inspired by the mathematical modeling of the intelligent behavior of spider monkeys, which follows a fission-fusion social structure (FFSS). According to FFSS, spider monkeys dynamically divide themselves from larger groups into smaller subgroups and vice versa to optimize their foraging activities. The main characteristics of FFSS are as follows [14] [27].

1) At the beginning, all spider monkeys form groups consisting of 40–50 individuals. Each group is guided by a leader, known as the global leader, who oversees the search for food sources.

2) If food resources are found to be insufficient, the global leader divides the larger group into smaller subgroups, each containing three to eight members, allowing them to forage independently. Each subgroup is led by a local leader.

3) The decision-making process for food searching within each subgroup is directed by a leader, referred to as the local leader.

4) Group members maintain social cohesion and establish defensive boundaries by communicating with one another and with members of other groups through a distinct vocalization.

In the Spider Monkey Optimization (SMO) algorithm, the initialization phase generates an initial population of N spider monkeys, distributed uniformly. Each individual spider monkey, denoted as $SM_i (i = 1, 2, \ldots, N)$, is represented as a vector of dimension D. Here, D signifies the number of variables in the optimization problem, and SM_i represents the i-th spider monkey in the population. Each spider monkey corresponds to a potential solution for the problem at hand [28]. The initialization of each SM_i is carried out according to the following equation:

$$SM_{ij} = SM_{minj} + U(0,1)\ x\ \left(SM_{maxj} - SM_{minj}\right) \quad (8)$$

Where $SM_{maxj}\ dan\ SM_{minj}$ are the bounds of SM_i in the j^{th} direction and $U(0,1)$ is a uniformly distributed random number with the range [0,1].

In the local leader stage of the Spyder Monkey Optimization (SMO) algorithm, each SM changes its current position by considering the information obtained from the experience of the local leader as well as the experience of the local group members [29]. The fitness value of the resulting new position is calculated. If the fitness value of the new position exceeds the fitness value of the old position, then the SM update its position to the new one. The equation used to update the position of the i-th SM (which is a member of the k-th local group) at this stage is as follows:

$$SMnew_{ij} = SM_{ij} + U(0,1)\ x\ \left(LL_{kj} - SM_{ij}\right) + U(-1,1)\ x\ \left(SM_{rj} - SM_{ij}\right) \quad (9)$$

where, SM_{ij} is the j-th dimension of the i-th SM, LL_{kj} represents the j-th dimension of the k-th local group leader position. SM_{rj} is the j^{th} dimension of the $SM\ r^{th}$ randomly chosen in the k^{th} group such that $r \neq 1, U(0,1)$ is a uniformly distributed random number between 0 and 1. After completing the local leader stage, the next step is the global leader stage. During this phase, all SM update their positions based on information from the global leader as well as the experiences of the local group members. The equation for updating positions in this stage can be formulated as follows:

$$SMnew_{ij} = SM_{ij} + U(0,1)\ x\ \left(GL_{kj} - SM_{ij}\right) + U(-1,1)\ x\ \left(SM_{rj} - SM_{ij}\right) \quad (10)$$

In this stage, where GL represents the j-th dimension of the global leader's position and $j \in \{1, 2, \ldots, D\}$ is a randomly chosen index, the positions of the spider monkeys (SM_i) are updated based on probabilities calculated from their fitness values. This approach ensures that candidates with better fitness values have a higher chance of improving their performance. The probability can be calculated using the following expression, which, although subject to variations, generally serves as a function of the fitness values:

$$prob_i = 0.9\ x\ \frac{fitness_i}{\max\_fitness} + 0.1 \quad (11)$$

Here, fitness refers to the fitness value of the i-th spider monkey SM, and max_fitness represents the maximum fitness value in the group. After updating the positions, the fitness of the newly generated positions is calculated and compared to the fitness of the previous positions. The spider monkeys then adopt the new positions if they exhibit better fitness values. This process involves recalculating and comparing the fitness

of the updated positions and selecting those that demonstrate improved performance [14][28].

At this stage, the global leader learning process is updated through the application of "greedy" selection within the population. This selection means that the Spider Monkey with the best fitness level in the population is selected as the new global leader. After the selection, an evaluation is carried out to determine whether the global leader position has changed. If there is no change, then the Global Limit Count value increased by one. Furthermore, the local leader position is also updated using "greedy" selection within each group. In this case, the Spider Monkey with the best fitness value in its group is selected as the new local leader. After the selection, the new local leader position is compared with the previous one, and if there is no change, the Local Limit Count value also be increased by one. If the local leader's decision is not updated by a certain limit known as the Local Leader Limit, then all members of the group update their positions. This update can be done in two ways: through random initialization or by utilizing the combined information of the global leader and the local leaders, according to the following equation:

$$SMnew_{ij} = SM_{ij} + U(0,1)\,x\,\left(GL_j - SM_{ij}\right) + U(0,1)\,x\,\left(SM_{ij} - LL_{kj}\right) \quad (12)$$

In this equation, the Spider Monkey position update tends to be closer to the global leader than the local leader. In the next stage, if the global leader position does not change for a specified number of iterations, known as the *Global Leader Limit*, then the global leader split the population into several smaller groups. Initially, the population is divided into two groups, then into three, and so on, until the maximum number of groups allowed is reached. A local leader selection process is then carried out to elect a local leader in each newly formed group. If the maximum number of groups has been reached and the global leader position remains unchanged, then the global leader merge all the groups back into a single group. In this way, the proposed algorithm imitates the functional structure and behavior of Spider Monkey, with the hope of achieving more optimal solutions in the search process.

D. Random Forest

Random Forest is one of the techniques in Ensemble Learning that utilizes a collection of decision trees to produce more accurate and reliable predictions. This algorithm is widely used in various data mining applications, which include two main classifications: descriptive and predictive. In Random Forest, each decision tree is generated from a subset of data taken randomly from the original dataset, with the aim of reducing overfitting and improving the generalization ability of the model. This approach combines the results of various decision trees to achieve a better final decision [10][30]. Mathematically, for a classification task, the final prediction $\hat{y}$ is defined as:

$$\hat{y} = \frac{1}{B}\sum_{b-1}^{B} h_b(x) \quad (13)$$

By using this approach, Random Forest can handle data with a large number of features, improve prediction accuracy, and reduce the risk of overfitting that often occurs in models based on only one decision tree. Its efficient implementation and reliable prediction results make Random Forest one of the methods widely used in various studies and practical applications in data mining [10][30].

E. Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a method developed by Vapnik for binary classification [31]. The primary objective of SVM is to find an optimal hyperplane, expressed as $f(w,x) = w.x + b$, that separates two classes in a given dataset with features $x \in Rm$. During the learning process, SVM determines the parameters www by solving an optimization problem as defined in Equation (5):

$$\min \frac{1}{p} W^T W + \mathrm{C}\sum_{i=1}^{p} max\left(0{,}1 - y_i'(W^T X_i + b)\right) \quad (14)$$

Here, $W^T W$ represents the Manhattan norm (also known as the L1 norm), and C is the penalty parameter, which can either be set to an arbitrary value or determined using hyper-parameter tuning. The term y' refers to the actual label, while $W^T X + b$ is the predictor function. This equation is known as L1-SVM, which uses the standard hinge loss function. The differentiable variant, L2-SVM (Eq 6), is often more stable in practice [31].

$$\min \frac{1}{p} \frac{\|W\|_2^2}{2} + \mathrm{C}\sum_{i=1}^{p} max\left(0{,}1 - y_i'(W^T X_i + b)\right)^2 \quad (15)$$

In this equation, $\|W\|2$denotes the Euclidean norm (also known as the L2 norm), and it uses the squared hinge loss. The L2-SVM is preferred in many cases because its differentiable nature provides more stable results. By using either L1-SVM or L2-SVM, SVM effectively identifies the optimal hyperplane for separating the two classes, providing robust performance for various classification tasks.

III. RESULT

The proposed method identifies diseases based on images using a state-of-the-art method consisting of four main phases. The first phase is Data Preprocessing, where the raw images are processed to remove noise, adjust the size, and separate relevant objects from the background. This process is essential to ensure optimal image quality before important features are extracted. In the second phase, key features of the image, such as color, texture, shape, and edges, are extracted to provide rich information about the characteristics of the described image. These features are then used as input for the classification process. The next phase is the third, the optimization and classification process is carried out. The Spyder Monkey Optimization (SMO) optimization method is used to improve the performance of the machine learning model, followed by the application of two classification algorithms, namely Random Forest and Support Vector Machine (SVM). The second algorithm is chosen because of its accuracy in handling complex data and providing reliable prediction results. Finally, the fourth phase is the evaluation of disease identification results, where the classification results are evaluated to assess the accuracy and effectiveness of the method used. This process aims to ensure that the applied method can accurately predict diseases based on the analyzed images (Fig. 3).

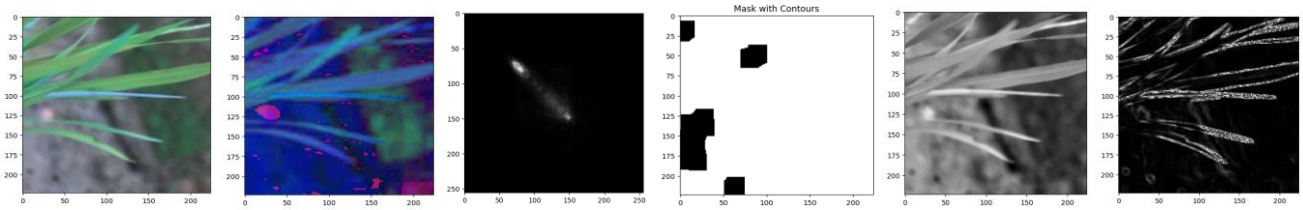


Fig. 3. Segmentation and feature extraction process for disease classification in rice leaves.

Fig. 4 on the top left and right, Random Forest (RF) produces very high Area Under the Curve (AUC), with some disease classes such as Tungro Virus and Grassy Stunt Virus reaching AUC = 1.00, indicating that the model is very accurate in distinguishing between positive and negative classes. Other diseases such as Bacterial Leaf Blight have slightly lower AUC, at AUC = 0.92. In the graphs on the bottom left, SVM with All Extraction Data shows relatively high performance, although there are some sharper fluctuations in FPR at the beginning of the curve. AUC for some classes, such as Rice Blast and Healthy Rice Plant, are in the range of AUC = 0.75 to 1.00, indicating that SVM also produces good results, but may not be as good as Random Forest in some cases. Finally, in the graph at the bottom right, SVM with Hyperparameter Tuning shows an increase in performance compared to SVM without tuning, with several disease classes having higher AUC values, approaching the maximum value of AUC = 1.00, indicating that parameter tuning has a significant impact on model performance.

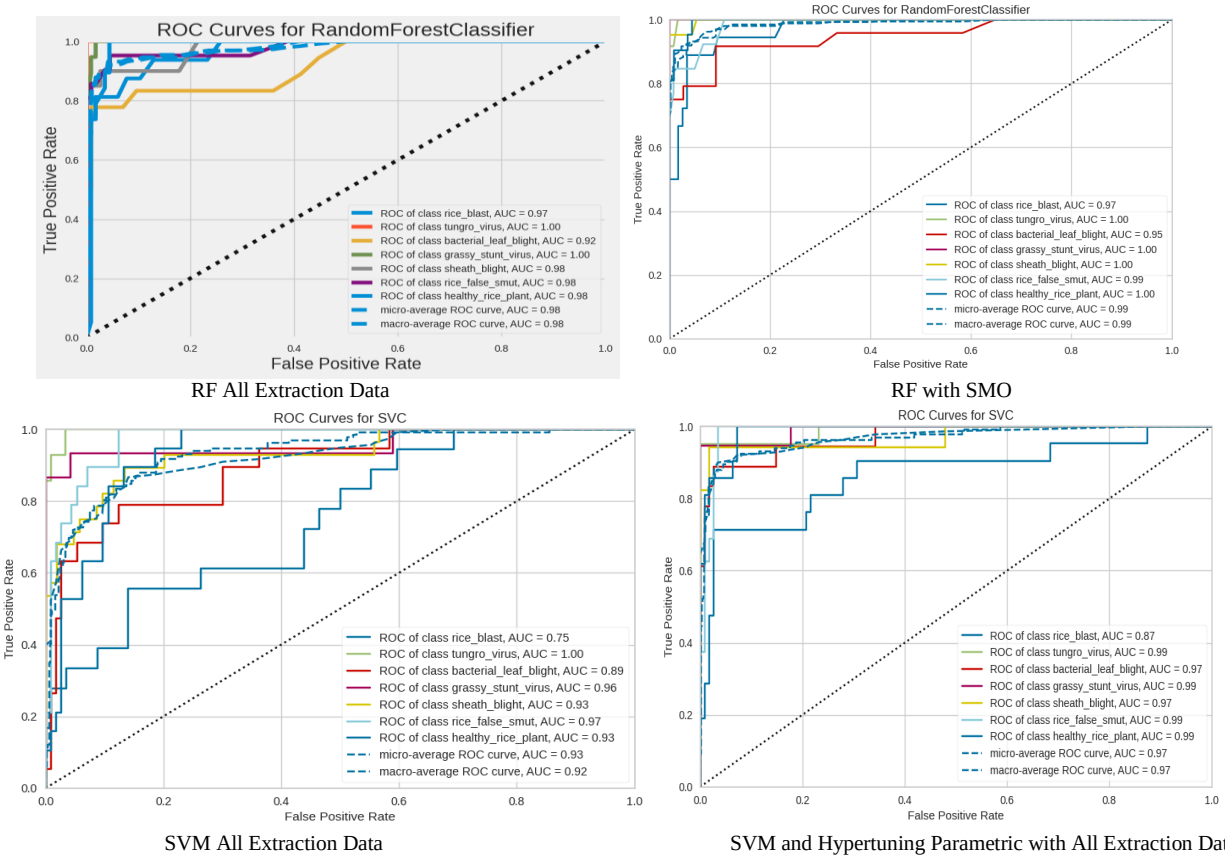


Fig. 4. The effect of using color and texture features in improving the performance of random forest and SVM (SMO) for classification of brown spot leaf disease in rice plants.

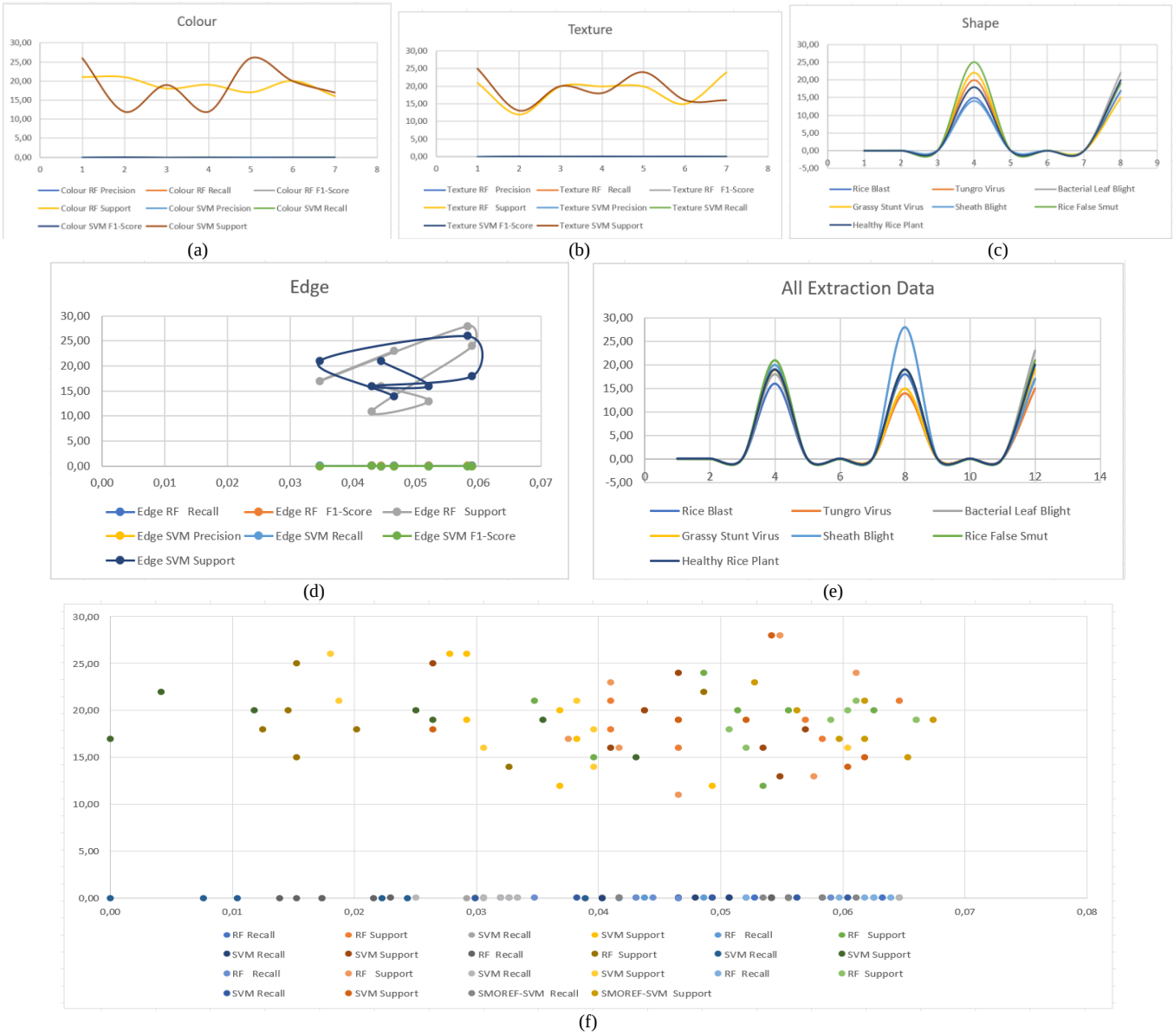


Fig. 5. Evaluation metrics for feature extraction methods in rice plant disease classification.

The given figure shows several graphs showing evaluation metrics such as Precision, Recall, F1 Score, and Support for various feature extraction methods (Color, Texture, Shape, and Edge) using classification algorithms such as Random Forest (RF), Support Vector Machine (SVM), and SMOREF-SVM. These graphs show the performance of various approaches in identifying rice plant diseases. From the results shown, it can be seen how the performance of these methods, both in individual feature processing and in combination, affects the model's ability to detect diseases accurately. In the Color feature extraction Fig. 5(a), the metrics for the RF and SVM algorithms show fluctuations between 10 and 25 on the y-axis. Precision and Recall for RF range between 10 and 25, with a peak between 3 and 5 on the x-axis. F1 Score follows a similar pattern. The SVM algorithm shows slightly lower performance, with metric values ranging from 10 to 20. From this, it can be seen that the Random Forest (RF) algorithm performs better than SVM in color feature extraction. The texture feature extraction in Fig. 5(b) shows a pattern similar to that of the color feature extraction. Precision for RF ranges from 10 to 25, with a peak around the 4th point on the x-axis. Recall for RF is slightly lower than Precision, but overall, it remains consistent. In the SVM algorithm, the Precision, Recall, and F1 Score metrics range from 10 to 20, which again shows that RF is superior to SVM in processing texture features. In the Shape feature extraction Fig. 5(c), the graph compares the performance of several rice diseases. Diseases such as Rice Blast, Tungro Virus, Bacterial Leaf Blight, and even Healthy Rice Plants all show peaks around the 25th value on the y-axis, with the highest point being around the 8th point on the x-axis. This shows that shape-based features provide very consistent results for various diseases, with very little variation in the evaluation metrics.

IV. DISCUSSION

Meanwhile, in the Edge feature extraction Fig. 5(d), the results show greater variation. Edge RF Recall ranges from 0.02 to 0.06 on the x-axis, and F1 Score ranges in the same range. However, the SVM algorithm in edge feature extraction has a lower Recall value, ranging from 0.01 to 0.04. This indicates that edge-based extraction does not provide consistent results, and the results achieved are lower than color, texture, or shape feature extraction, with RF tending to perform better than SVM. Fig illustrated combining data from all extraction methods Fig. 5(e) shows a pattern of peak values for diseases such as Rice Blast, Tungro Virus, and Bacterial Leaf Blight, all of which are in the range of 20 to 25. This graph shows that when the extraction methods are combined, the evaluation metrics become more stable and show reliable performance for various rice diseases. In the scatter plot combining the RF, SVM, and SMOREF-SVM algorithms Fig. 5(f), the Recall and Support metrics for each algorithm are shown. RF Recall mostly clusters between 0.02 and 0.06 on the x-axis, with Support values varying between 15 and 25. The SVM algorithm shows a more spread-out distribution, with some points falling below 0.02. Meanwhile, SMOREF-SVM shows a similar clustering pattern to RF, although there are some lower variations on the x-axis (Table II).

TABLE II. COMPARISON OF THE PERFORMANCE OF THE PROPOSED METHOD WITH PREVIOUS STUDIES

Similarities Research	Year	Methods Used	Other Similarities	Result Performance
[7]	2019	- Image Acquisition, Pre-processing, Segmentation, Classification - Hybrid method: Discrete Wavelet Transform (DWT), Scale Invariant Feature Transform (SIFT), Grayscale Co-occurrence Matrix (GLCM) - Classifiers: KNN, BPNN, Naïve Bayes, Multiclass SVM	- Focus on rice disease detection using image processing and classification.	98.63%
[8]	2019	- Image Acquisition - Preprocessing (Grayscale conversion, segmentation) - Neural Network (Pattern Recognition)	- Detection of rice diseases with artificial neural networks.	92.5%
[9]	2021	- Image Acquisition, Pre-processing, Thresholding, Edge Detection, Color Slicing - RBG Calculation for Classification	- Focus on color techniques to detect rice leaf diseases, especially Brown Spot and Narrow Brown Spot.	89%
[10]	2021	- Image Acquisition - Preprocessing - Random Forest Decision Tree Classifier - Feature extraction with Intensity Moments	- Detection of rice diseases using Random Forest with classification of three main types of diseases (Blight, Blast, Spot).	91.47%
Proposed method	2024	- Spider Monkey Optimization (SMO) for Feature Optimization - Random Forest (RF) for Initial Classification - Support Vector Machine (SVM) for Final Classification - ROC Curve Analysis - Performance Metrics: Precision, Recall, F1 Score	- Combination of Random Forest and SVM with feature optimization using SMO. Focus on improving disease classification accuracy.	98% (AUC = 0.98, Precision = 94%, Recall = 92%, F1 Score = 93%)

This study offers a more comprehensive picture of model performance compared to [7]. While both models have high accuracy rates—with [7] reaching 98.63% and this study reaching 98%—this study stands out in providing more in-depth evaluation metrics, such as Precision, Recall, and F1 Score. With Precision reaching 94%, Recall 92%, and F1 Score 93%, this study provides a more balanced view of the model's ability to consistently classify data and capture relevant patterns. Coupled with an AUC approaching 0.98, the SMOREF-SVM model used in this study not only excels in terms of accuracy but is also more effective in handling datasets that may have uneven class distributions. In contrast, while [7] shows high accuracy rates, the lack of additional metrics such as Precision, Recall, and F1 Score makes it difficult to comprehensively evaluate the model's performance across situations. Therefore, this study excels in providing a holistic assessment of model performance, especially in the context of real-world use cases where successful predictions on the minority class are as important as overall accuracy.

V. CONCLUSIONS

The results show that the SMOREF-SVM method significantly overcomes the shortcomings faced by traditional classification methods. Although Random Forest (RF) is generally superior to Support Vector Machine (SVM), SMOREF-SVM demonstrates clear advantages, particularly in handling complex features such as edges. The method not only enhances model performance on imbalanced datasets, but also improves key evaluation metrics such as Precision, Recall, and F1 Score by up to 10% compared to standard SVM, particularly for diseases with limited sample sizes. Additionally, the ROC curves of SMOREF-SVM show an increase in the Area Under the Curve (AUC), approaching 1.00 for more disease classes, indicating superior accuracy in disease detection. However, there are still several areas that require further investigation. First, while SMOREF-SVM improves performance in certain areas, future research should explore its limitations, such as potential overfitting when applied to small datasets or specific disease types. Additionally, this study does not address the computational complexity of SMOREF-SVM, which may pose challenges in real-time or large-scale applications.

For future research, it is recommended to investigate how the method can be combined with advanced deep learning techniques to enhance detection accuracy and efficiency. Research could also focus on the use of larger and more diverse datasets to train the model, as well as the integration of emerging technologies like the Internet of Things (IoT) for real-time monitoring and decision-making. Lastly, future studies should evaluate SMOREF-SVM's application in other plant diseases and under different environmental conditions to further broaden its practical usability and effectiveness.

REFERENCES

[1] C. Sarkar, D. Gupta, U. Gupta, and B. B. Hazarika, "Leaf disease detection using machine learning and deep learning: Review and challenges," Appl. Soft Comput., vol. 145, p. 110534, 2023, doi: 10.1016/j.asoc.2023.110534.

[2] S. S. Harakannanavar, J. M. Rudagi, V. I. Puranikmath, A. Siddiqua, and R. Pramodhini, "Plant leaf disease detection using computer vision and machine learning algorithms," Glob. Transitions Proc., vol. 3, no. 1, pp. 305–310, 2022, doi: https://doi.org/10.1016/j.gltp.2022.03.016.

[3] K. Effendi, MunifAbdul, and W. Wayan, "Pengetahuan, Sikap, dan Tindakan Petani Upsus dalam Mengendalikan Hama dan Penyakit Tanaman Padi," J. Ilmu Pertan. Indones., vol. 25, no. 4, pp. 515–523, 2020, doi: 10.18343/jipi.25.4.515.

[4] L. Budiarti, J. Kartahadimadja, M. Sari, D. Ahyuni, and D. Dulbari, "Keanekaragaman Artropoda Predator di Agroekosistem Sawah pada Berbagai Galur Padi Politeknik Negeri Lampung," AGROSCRIPT J. Appl. Agric. Sci., vol. 3, 2021, doi: 10.36423/agroscript.v3i1.663.

[5] P. Kartikeyan and G. Shrivastava, "Review on Emerging Trends in Detection of Plant Diseases using Image Processing with Machine Learning," Int. J. Comput. Appl., vol. 174, no. 11, pp. 39–48, 2021, doi: 10.5120/ijca2021920990.

[6] T. Bera, A. Das, J. Sil, and A. K. Das, "A Survey on Rice Plant Disease Identification Using Image Processing and Data Mining Techniques," in Emerging Technologies in Data Mining and Information Security, A. Abraham, P. Dutta, J. K. Mandal, A. Bhattacharya, and S. Dutta, Eds., Singapore: Springer Singapore, 2019, pp. 365–376.

[7] T. Gayathri Devi and P. Neelamegam, "Image processing based rice plant leaves diseases in Thanjavur, Tamilnadu," Cluster Comput., vol. 22, no. 6, pp. 13415–13428, 2019, doi: 10.1007/s10586-018-1949-x.

[8] M. Biswas, N. Faruqui, M. Lata, L. Lecturer, M. Julkar, and M. J. Nayeen Mahi, "A Novel Inspection of Paddy Leaf Disease Classification using Advance Image Processing Techniques," 2019.

[9] N. S. A. M. Taujuddin, I. S. Samuri, R. Ibrahim, S. Sari, A. R. A. Ghani, and M. A. Rahman, "Brown Spot and Narrow Brown Spot Paddy Disease Detection using Color Slicing Method," Evol. Electr. Electron. Eng., vol. 2, no. 1, pp. 45–54, 2021, [Online]. Available: http://publisher.uthm.edu.my/periodicals/index.php/eeee

[10] S. Saha and S. M. M. Ahsan, "Rice Disease Detection using Intensity Moments and Random Forest," in 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), 2021, pp. 166–170. doi: 10.1109/ICICT4SD50815.2021.9396986.

[11] P. Pavithra and P. Aishwarya, "Plant leaf disease detection using hybrid grasshopper optimization with modified artificial bee colony algorithm," Multimed. Tools Appl., vol. 83, no. 8, pp. 22521–22543, 2024, doi: 10.1007/s11042-023-16148-5.

[12] V. Singh, "Sunflower leaf diseases detection using image segmentation based on particle swarm optimization," Artif. Intell. Agric., vol. 3, pp. 62–68, 2019, doi: https://doi.org/10.1016/j.aiia.2019.09.002.

[13] J. Zhao, Y. Fang, G. Chu, H. Yan, L. Hu, and L. Huang, "Identification of Leaf-Scale Wheat Powdery Mildew (Blumeria graminis f. sp. Tritici) Combining Hyperspectral Imaging and an SVM Classifier," Plants, vol. 9, no. 8, 2020, doi: 10.3390/plants9080936.

[14] S. Kumar, B. Sharma, V. K. Sharma, H. Sharma, and J. C. Bansal, "Plant leaf disease identification using exponential spider monkey optimization," Sustain. Comput. Informatics Syst., vol. 28, p. 100283, 2020, doi: 10.1016/j.suscom.2018.10.004.

[15] N. S. Patil, "Identification of Paddy Leaf Diseases using Evolutionary and Machine Learning Methods," Turkish J. Comput. Math. Educ., vol. 12, no. 2, pp. 1672–1686, 2021, doi: 10.17762/turcomat.v12i2.1503.

[16] T. A. Lestari, "Pengamatan Penyakit-Penyakit Tanaman Sistem Jajar Legowo Observation of Paddy Diseases in The Vilage of Sako Rambutan Sub-District Banyuasin with The Jajar Legowo System," J. Ilm. Sinus, 2019.

[17] A. Walascha et al., "Review Artikel: Inventarisasi Jenis Penyakit yang Menyerang Daun Tanaman Padi (Oryza sativa L.)," Pros. Semin. Nas. Biol., vol. 1, no. 2, pp. 471–478, 2021, [Online]. Available: https://semnas.biologi.fmipa.unp.ac.id/index.php/prosiding/article/download/150/300

[18] H. Herdiyanti, Eko Sulistyono, and Purwono, "Pertumbuhan dan Produksi Beberapa Varietas Padi (Oryza sativa L.) pada Berbagai Interval Irigasi," J. Agron. Indones. (Indonesian J. Agron., vol. 49, no. 2, pp. 129–135, 2021, doi: 10.24831/jai.v49i2.36558.

[19] Petchiammal, B. Kiruba, Murugan, and P. Arjunan, "Paddy Doctor: A Visual Image Dataset for Automated Paddy Disease Classification and Benchmarking," in Proceedings of the 6th Joint International Conference on Data Science \& Management of Data (10th ACM IKDD CODS and 28th COMAD), in CODS-COMAD '23. New York, NY, USA: Association for Computing Machinery, 2023, pp. 203–207. doi: 10.1145/3570991.3570994.

[20] S. Lamba, A. Baliyan, and V. Kukreja, "A novel GCL hybrid classification model for paddy diseases," Int. J. Inf. Technol., vol. 15, no. 2, pp. 1127–1136, 2023, doi: 10.1007/s41870-022-01094-6.

[21] M. A. Azim, M. K. Islam, M. M. Rahman, and F. Jahan, "An effective feature extraction method for rice leaf disease classification," Telkomnika (Telecommunication Comput. Electron. Control., vol. 19, no. 2, pp. 463–470, 2021, doi: 10.12928/TELKOMNIKA.v19i2.16488.

[22] S. Sakhamuri and K. K. Kumar, "Deep Learning and Metaheuristic Algorithm for Effective Classification and Recognition of Paddy Leaf Diseases," J. Theor. Appl. Inf. Technol., vol. 100, no. 4, pp. 1127–1137, 2022.

[23] S. Neware, "Paddy plant leaf diseases identification using machine learning approach," Int. J. Health Sci. (Qassim)., vol. 6, no. S1, pp. 10467–10472, 2022, doi: 10.53730/ijhs.v6nS1.7522.

[24] V. K. Vishnoi, K. Kumar, and B. Kumar, "Plant disease detection using computational intelligence and image processing," J. Plant Dis. Prot., vol. 128, no. 1, pp. 19–53, 2021, doi: 10.1007/s41348-020-00368-0.

[25] K. K. Thyagharajan and I. Kiruba Raji, "A Review of Visual Descriptors and Classification Techniques Used in Leaf Species Identification," Arch. Comput. Methods Eng., vol. 26, no. 4, pp. 933–960, 2019, doi: 10.1007/s11831-018-9266-3.

[26] M. Sharif, M. A. Khan, Z. Iqbal, M. F. Azam, M. I. U. Lali, and M. Y. Javed, "Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection," Comput. Electron. Agric., vol. 150, pp. 220–234, 2018, doi: https://doi.org/10.1016/j.compag.2018.04.023.

[27] J. C. Bansal, H. Sharma, S. S. Jadon, and M. Clerc, "Spider Monkey Optimization algorithm for numerical optimization," Memetic Comput., vol. 6, no. 1, pp. 31–47, 2014, doi: 10.1007/s12293-013-0128-0.

[28] D. Kumari, A. Sinha, S. Dutta, and P. Pranav, "Optimizing neural networks using spider monkey optimization algorithm for intrusion detection system," Sci. Rep., vol. 14, no. 1, pp. 1–16, 2024, doi: 10.1038/s41598-024-68342-6.

[29] A. Agrawal, D. Garg, R. Sethi, and A. K. Shrivastava, "Optimum redundancy allocation using spider monkey optimization," Soft Comput., vol. 27, no. 21, pp. 15595–15608, 2023, doi: 10.1007/s00500-023-08746-0.

[30] V. Kulkarni and P. Sinha, "Random forest classifiers: A survey and future research directions," Int. J. Adv. Comput., vol. 36, pp. 1144–1153, Jan. 2013.

[31] A. F. Agarap, "An Architecture Combining Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for Image Classification," pp. 5–8, 2017, [Online]. Available: http://arxiv.org/abs/1712.03541.

Human Dorsal Hand Vein Segmentation Method Based on GR-UNet Model

Zhike Zhao[1], Wen Zeng[2], Kunkun Wu[3], Xiaocan Cui[4]
School of Electrical Engineering, Henan University of Technology, Zhengzhou, China[1, 2]
Henan Baichangyuan Medical Technology Co., LTD, Zhengzhou, China[3]
Xinxiang First People's Hospital, Xinxiang, China[4]

***Abstract*—To solve the issue of inaccurate segmentation accuracy of human dorsal hand veins (HDHV), we propose a segmentation method based on the global residual U-Net (GR-Unet) model. Initially, a visual acquisition device for dorsal hand vein imaging was designed utilizing near-infrared technology, resulting in the creation of a dataset comprising 864 images of HDHV. Subsequently, a Bottleneck from the deep residual network-50 (ResNet50) was integrated into the U-Net model to enhance its depth and alleviate the problem of vanishing gradients. Furthermore, a global attention mechanism (GAM) was introduced at the junction to improve the acquisition of global feature information. Additionally, a weighted loss function that combines cross-entropy loss and Dice loss was employed to address the imbalance between positive and negative samples. The experimental results indicate that the GR-Unet model achieved accuracies of 78.82%, 88.03%, 93.92%, and 97.5% in terms of intersection over union, mean intersection over union, mean pixel accuracy, and overall accuracy, respectively.**

***Keywords*—Human dorsal hand veins; GR-UNet; near infrared technology; deep residual network-50; global attention mechanism; loss function**

I. INTRODUCTION

Venipuncture is a critical procedure for blood collection, transfusion, and infusion in clinical medicine, with the hand serving as the primary site for this practice. The conventional technique for venipuncture on the dorsal aspect of the hand involves direct venipuncture, during which the patient clenches their fist and utilizes a pressure band to engorge the vein. Subsequently, medical personnel assess the optimal site for venipuncture based on clinical experience before performing the procedure. However, due to variations in the distribution of human veins, achieving accurate venous puncture through manual means alone can be challenging, particularly in children, the elderly, and obese individuals. These populations often present with thinner venous vessels, poorer vascular elasticity, and increased perivenous fat, which complicates venous positioning in the context of venipuncture [1]. Therefore, enhancing the imaging of hand veins using computer-assisted methods may be crucial for improving venous positioning.

Venous positioning is a prerequisite for successful venipuncture. Numerous researchers have conducted extensive studies to accurately locate venous vessels. Ma et al. [2] processed images of the dorsal hand vein using contrast-limited adaptive histogram equalization and multi-scale detail fusion algorithms, subsequently weighting and superimposing the two to enhance the vein images. Although this method addressed the issue of detail loss associated with the histogram equalization algorithm, it remained ineffective in managing vein images that contained high-frequency noise. Kuang et al. [3] employed multi-scale Gaussian blurring to denoise the L component extracted from the LAB color space representation of the image, and utilized guided filtering for illuminance estimation, combined with adaptive thresholding for dynamic adjustment to enhance the vein image. However, the algorithm's adaptability to various application environments required further improvement. Besra et al. [4] proposed a vein segmentation algorithm based on the repeated line tracking method, which achieved vein segmentation by superimposing the trajectory lines in the vein images tracked from different starting points. Nonetheless, this method was highly dependent on image clarity and involved a substantial number of arithmetic operations for image processing, making it challenging to meet the demands of real-time detection and recognition. Yakno et al. [5] employed contrast-limited adaptive histogram equalization and fuzzy adaptive gamma transform to enhance vein image processing. They subsequently weighted the two methods and applied a matched filter with a first-order derivative of the Gaussian function to achieve vein image segmentation. However, this approach required multiple transformations and integrations of the image, leading to high computational complexity and limited effectiveness in managing high-brightness areas. In contrast, Yang et al. [6] utilized a six-dimensional Gabor filter for vein image enhancement and a Markov random field for vein image segmentation. Nonetheless, this method demonstrated poor performance in handling low-contrast vein images or those containing high-frequency noise.

With the rapid development of deep learning theory, achieving vein image segmentation using convolutional neural networks has become increasingly feasible [7]. In 2015, Long et al. [8] proposed Fully Convolutional Networks, which processed pixel-level data by employing full convolution instead of full connectivity, thereby allowing for the processing of image inputs of any size. However, this approach did not adequately address the relationships between pixels when processing them independently, leading to incomplete feature information extraction. In the same year, Ronneberger et al. [9] introduced the U-Net network algorithm for image segmentation, which effectively addressed challenges in low-resolution image processing and has since gained widespread application in the field of medical image segmentation. The direct application of the U-Net network model to venous vessel

segmentation still presents several limitations: (1) A disproportionately small sample set leads to overfitting during network training; (2) The single encode-decode structure may cause the network training to reach a performance ceiling, making it challenging to extract additional effective features; (3) Although the U-Net decoding stage integrates local feature information through skip connections, simple concatenation does not fully merge features of varying scales. To address these issues, researchers have proposed enhanced algorithms based on the U-Net model by modifying the backbone and incorporating an attention mechanism. He et al. [10] introduced an algorithm that leverages the U-Net model alongside the attention mechanism, which improved the identification of multi-scale features and enhanced segmentation accuracy, although it was less effective in segmenting narrow venous vessels. Lefkovits et al. [11] proposed a hybrid approach combining unsupervised and supervised techniques for dorsal hand vein segmentation, experimentally demonstrating that the segmentation results of eight U-Net variants surpassed those of traditional image methods, with the ResNet-U-Net model exhibiting particularly significant performance. Gao et al. [12] proposed a semantic segmentation model called AT-U-Net, which integrates the Non-Local attention mechanism into the U-Net architecture to enhance feature extraction capabilities. This approach addresses the challenge of long-distance dorsal hand vein puncture; however, the limitation in capturing venous vessel images from afar may result in inadequate extraction of global feature information. Additionally, high-frequency noise and blurred vessel edges can adversely affect segmentation accuracy. Chen et al. [13] incorporated a Gabor convolution kernel into the U-Net framework and utilized inverted residual blocks to achieve model lightweighting, thereby mitigating the semantic information loss attributed to channel issues. Nonetheless, this algorithm depends heavily on the extraction of shallow features, and the parameters of the Gabor filter require manual adjustment, which adds to the algorithm's complexity.

In contrast to the limitations of traditional image segmentation algorithms, neural networks are more adept at managing complex image features and capturing nonlinear relationships within images, thereby achieving enhanced segmentation accuracy and robustness. Consequently, to improve the extraction of HDHV, this paper proposes a segmentation algorithm based on GR-Unet. The Bottleneck portion of the ResNet50 serves as the backbone network, while GAM is integrated into the skip connections, along with pre-trained weights to expedite feature extraction. The network's capacity to learn from defective pixels is further augmented through a weighted loss function that combines cross-entropy and Dice loss. Experiments conducted on a self-constructed dorsal hand vein dataset validate the effectiveness of the proposed method.

II. Design Scheme

The design scheme of this paper encompasses several key components: image acquisition, image preprocessing, data enhancement, dataset construction, GR-UNet model development, network training, and experimental testing sessions, as illustrated in Fig. 1. Specifically, image preprocessing involved grayscale normalization, global threshold binarization, the centroid method for extracting regions of interest (ROI), averaging filtering, and contrast-limited adaptive histogram equalization (CLAHE). The GR-UNet model utilized ResNet50 as its backbone network, incorporating the GAM module and a modified loss function at the skip connections.

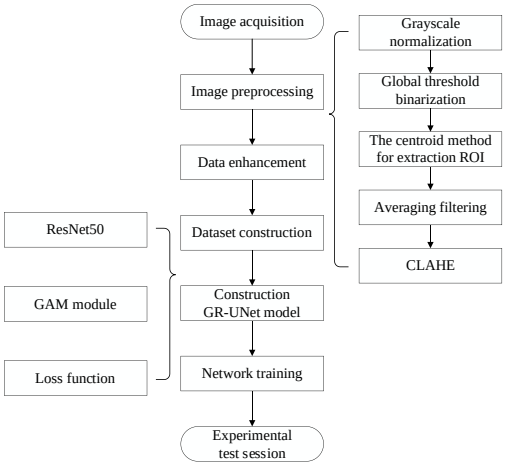


Fig. 1. The technical roadmap of this study.

III. Dorsal Hand Vein Dataset

A. Dorsal Hand Vein Collection System Design

Due to the differential absorption of near-infrared (NIR) light by human tissues, hemoglobin in human veins can absorb NIR light in the range of 700 nm to 1100 nm [14]. Consequently, images within this spectral band (700 nm to 1100 nm) can be captured using a camera equipped with a suitable filter. However, during the experimental process, results indicated that while NIR light could penetrate 5 to 10 mm into human tissues [14], it was less effective in capturing images from individuals with higher fat content in their hands. Existing studies have demonstrated that the use of 850 nm NIR light can effectively highlight the contour structure of veins [15-16], whereas adipose tissue exhibits optimal absorption characteristics at the 940 nm NIR wavelength band [17]. Therefore, this paper proposes a dual-band measurement superposition method, which involves utilizing a 940 nm NIR light source for top irradiation combined with an 850 nm NIR light source for bottom irradiation to facilitate the acquisition of human hand vein images. The schematic representation of the designed human hand vein image acquisition system is illustrated in Fig. 2.

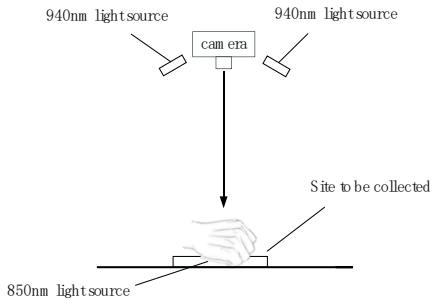


Fig. 2. Schematic diagram of human hand back image acquisition equipment.

In this study, a total of 72 male and female subjects were experimentally recruited, with ages ranging from 3 to 60 years. Among these participants, 30 were minors and 42 were adults. Each subject had images captured of their left and right hand veins, resulting in a total of 144 images, each with a resolution of 640x480 pixels. To enhance the diversity of the image samples, an additional 864 images of human hand veins were generated through rotation and mirroring of the collected images. The images of human dorsal hand vein obtained after NIR filtering is shown in Fig. 3.

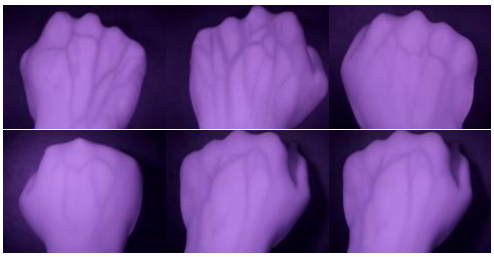


Fig. 3. The images of the human dorsal hand vein obtained after NIR filtering.

B. Image Preprocessing

Due to the interference of realistic conditions, including variations in illumination and the differing thickness of the dorsal hand vein among individuals, the contrast of the collected dorsal hand vein images was suboptimal and contained various artifacts. This compromised the clarity of the boundary between the veins and the background, leading to potential misjudgment in identifying the veins. Consequently, to accurately segment the venous vessels on the dorsal hand vein, it is essential to preprocess the vein images to effectively correct for uneven illumination and eliminate shadows and artifacts. The specific steps of image preprocessing are as follows:

Step 1: Grayscale normalization was employed to standardize the grayscale values of the dorsal hand vein image to a range of 0 to 255, with the calculation formula presented in Eq. (1).

$$P(x,y)=\frac{\left[R(x,y)-R_{\min}(x,y)\right]\times 255}{R_{\max}(x,y)-R_{\min}(x,y)} \tag{1}$$

Wherein, $R(x,y)$ denotes the gray value of the dorsal hand vein image, $R_{\min}(x,y)$ and $R_{\max}(x,y)$ denote the minimum and maximum values of the gray value of the dorsal hand vein image, respectively, and $P(x,y)$ denotes the normalized gray value of the dorsal hand vein image.

Step 2: The dorsal hand vein images were binarized using the global thresholding method, resulting in the dorsal part of the hand appearing in white against a black background.

Step 3: The centroid method $G(x_i, y_j)$ was applied to the binary image to calculate the centroid, and a ROI measuring 288×288 pixels was extracted using this centroid as the reference point. The formulas for the centroid method calculations are provided in Eq. (2) and Eq. (3).

$$x_i=\frac{\sum_{i=0}^{h}\sum_{j=0}^{w}\left[i\times f(i,j)\right]}{\sum_{i=0}^{h}\sum_{j=0}^{w}f(i,j)} \tag{1}$$

$$y_i=\frac{\sum_{i=0}^{h}\sum_{j=0}^{w}\left[j\times f(i,j)\right]}{\sum_{i=0}^{h}\sum_{j=0}^{w}f(i,j)} \tag{3}$$

Where, h and w denote the height and width of the image, respectively, and $f(x,y)$ denotes the pixel point in the white area.

Step 4: An averaging filter was applied to the ROI to reduce the Gaussian noise present in the image.

Step 5: The CLAHE algorithm was utilized to enhance the contrast of the ROI. Finally, the processed image of the dorsal hand vein is presented in Fig. 4.

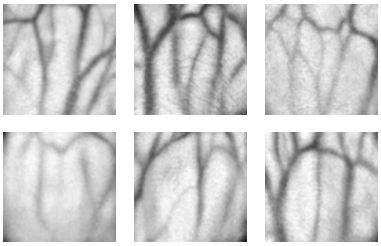


Fig. 4. Preprocessed image results of dorsal hand vein.

C. Dataset Construction

The preprocessed images of HDHV were labeled on a pixel-by-pixel basis, identifying the dorsal hand vein vessels. This process established the labeled images and ultimately constructed the dataset. Eighty percent of the images were randomly assigned to the training set, while ten percent were designated for the validation set and the remaining ten percent for the test set. Randomly selected original and labeled images from the dataset are presented in Fig. 5.

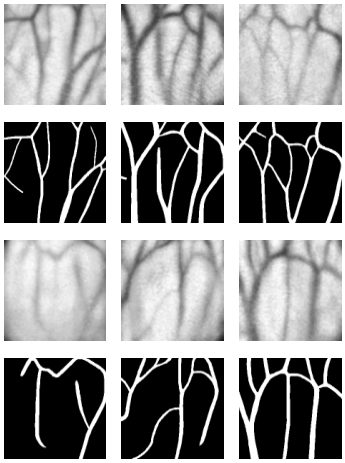


Fig. 5. Original and labeled images.

IV. METHODS

A. Basic Network Model

The U-Net model, developed by Ronneberger et al. [9], was initially applied as a convolutional neural network for cell segmentation in medical images. It features a U-shaped symmetric structure divided into encoding and decoding networks, as illustrated in Fig. 6. The left side comprises the encoding network, which consists of four subsampled modules. Each module contains two consecutive 3×3 convolutional layers, the ReLU activation function, and a 2×2 max-pooling layer. The primary purpose of the encoding network is to extract information from shallow features, as the size of the feature map decreases while the number of feature channels increases. Conversely, the right side contains the decoding network, which is upsampled by stacking 2×2 transposed convolutional layers. After each upsampling, it is spliced with the corresponding feature layer from the encoding network using skip connections, thereby fusing feature information from both deeper and shallower layers. This combined information is then processed by convolutional blocks to gradually recover the size and spatial dimensions of the input image. Ultimately, the segmentation result is obtained through a 1×1 convolutional layer.

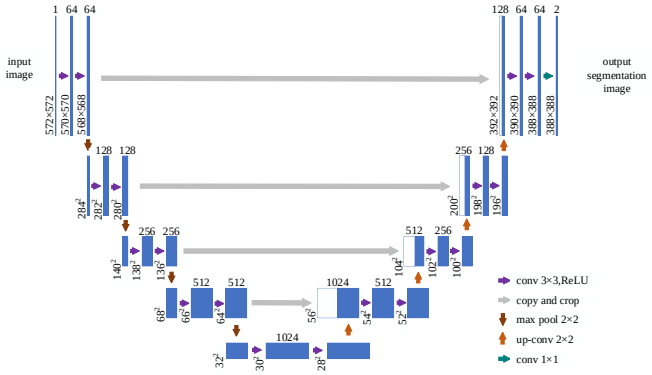


Fig. 6. U-Net model structure.

This paper proposes a GR-UNet model structure divided into two parts: encoding and decoding, as illustrated in Fig. 7. The encoding component utilizes the bottleneck part of the ResNet50 network, effectively mitigating issues related to vanishing gradients and network degradation. In the decoding phase, five up-sampling operations are conducted using bilinear interpolation, followed by two 3×3 convolutions with ReLU activation functions after each up-sampling to eliminate the aliasing effects present in the feature maps. Additionally, the GAM attention mechanism is incorporated at the skip connections to enhance the extraction of global feature information. This allows the network to focus more on the features of HDHV across both spatial and channel dimensions, thereby assigning greater weight coefficients within the network.

B. The ResNet50 Network

To obtain deeper features, this paper replaces the encoding component of the U-Net network with the Bottleneck module from the ResNet50 architecture. The global average pooling layer and the fully connected layer are removed, resulting in the construction of the ResNet50-UNet network. This modified architecture is employed to perform the vein segmentation task. The ResNet50-UNet network effectively combines the strengths of both the U-Net and ResNet50 architectures, thereby mitigating issues related to vanishing gradients and network degradation. Additionally, it fully leverages shallow features to enhance the efficiency of feature extraction when analyzing HDHV.

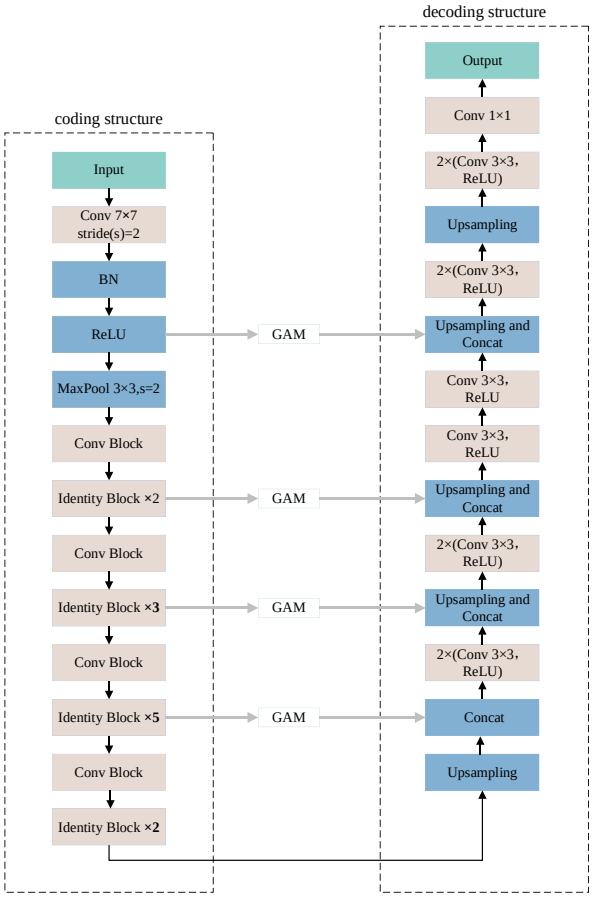


Fig. 7. GR-UNet model structure.

The ResNet50 network was proposed by He et al. [18] and demonstrates excellent performance across various visual application domains, including classification, object detection, and semantic segmentation. The architecture of the ResNet50 network primarily consists of an input layer, a residual block, a global average pooling layer, and a fully connected layer, as illustrated in Fig. 8. The residual mapping component on the left side of the Conv Block undergoes three convolution processes: a 1×1 convolution (with a stride of 2), a 3×3 convolution (with a stride of 1), and another 1×1 convolution (with a stride of 1). Following each convolution, a Batch Normalization (BN) layer is employed to normalize the features. The BN layer associated with the first two convolutions is succeeded by a ReLU activation function, which helps mitigate the vanishing gradient problem. In the right direct mapping component, a 1×1 convolution (with a stride of 2) and a BN layer normalization are performed, after

which the feature map output from the residual mapping is added pixel-by-pixel to the feature map output from the direct mapping. The combined output is then processed through a ReLU activation function, resulting in a feature map that is half the size of the input feature map. The Identity Block is analogous to the Conv Block; however, the left side maintains the same structure as the Conv Block, with all convolutional layers having a stride of 1. The right side features a direct mapping that does not include the 1×1 convolution operation or BN layer normalization, allowing the inputs to be added directly to the outputs, which subsequently pass through the ReLU activation function. The Conv Block is designed to reduce the size of the feature map, while the Identity Block serves to deepen the network.

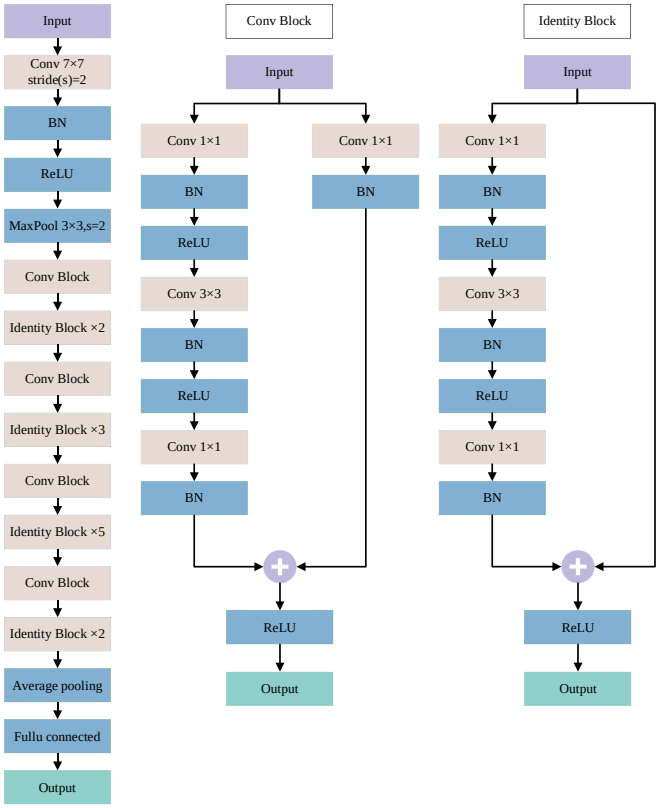


Fig. 8. ResNet50 network structure.

C. The GAM Attention Module

The decoder construction of U-Net closely resembles that of the encoder. Popular upsampling operations, such as transpose convolution, are local operations [19]. However, they do not adequately mine the in-depth feature information of the context, leading to incomplete information being conveyed during skip connections and subsampling. Consequently, this results in a failure to obtain rich feature information. To address this issue and achieve a more comprehensive understanding of HDHV, the GAM attention module has been introduced at the skip junction. This enhancement facilitates the acquisition of global feature information, allowing for further optimization and processing of deep feature information to better capture the global contextual information. The GAM attention mechanism is designed to minimize information loss while amplifying global dimensional interaction features [20]. It is a redesign based on the CBAM [21] attention mechanism, which also employs spatial and channel attention mechanisms. Its structure is illustrated in Fig. 9.

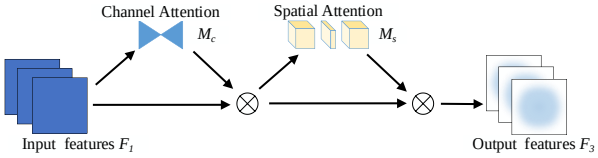


Fig. 9. GAM attention module.

In the whole process, the input features are $F_1 \in R^{C\times H\times W}$, F_2 is the intermediate state, and F_3 is the output features, and the relationship between them is shown in Eq. (4) and Eq. (5):

$$F_2 = M_C(F_1) \otimes F_1 \tag{2}$$

$$F_3 = M_s(F_2) \otimes F_2 \tag{3}$$

Where, M_C and M_s are channel feature maps and spatial feature maps respectively, and $\otimes$ is the multiplication operation.

In the channel attention submodule, a three-dimensional arrangement is initially applied to the input data to preserve its information across three distinct dimensions. Subsequently, the cross-dimensional channel and spatial dependencies are enhanced by a two-layer Multilayer Perceptron (MLP), which performs a dimensionality transformation, reverting the dimensions to their original state. Finally, the output is processed through the Sigmoid activation function, allowing for the fusion of information across different dimensions by implementing the dimensional transformation. The structure of the channel attention submodule is illustrated in Fig. 10.

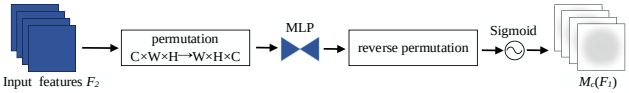


Fig. 10. Channel attention submodule.

In the spatial attention submodule, the convolution operation is primarily performed on the input feature graph F_2. Initially, information from the spatial layers is fused using two 7×7 convolutional layers to enhance the learning of spatial features. The output is then processed through the Sigmoid activation function to enhance the integration of data across different dimensions. The structure of the spatial attention submodule is illustrated in Fig. 11.

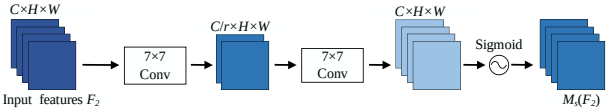


Fig. 11. Spatial attention submodule.

D. Improved Loss Function

To address the imbalance between positive and negative samples, this paper improves the original loss function by employing a weighted combination of Cross-Entropy Loss (CE

Loss) and Dice Loss. The calculation of the loss function is detailed in Eq. (6) to Eq. (8).

$$Loss = CE\_Loss + Dice\_Loss \tag{6}$$

$$CE\_Loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log\left(p_{i,c}\right) \tag{7}$$

$$Dice\_Loss = 1 - \frac{2\times\left(\sum_{i=1}^{N} P_i y_i\right)}{\sum_{i=1}^{N} P_i + \sum_{i=1}^{N} y_i} \tag{8}$$

Where CE_Loss is the cross-entropy loss function, $Dice\_Loss$ is the Dice loss function, N is the number of samples, C is the number of classes, $y_{i,c}$ is the actual label of sample i belonging to class c, $p_{i,c}$ is the probability of sample i belonging to class c, P_i and y_i are the predicted value of the model and the real label respectively.

V. Experiments

A. The Experimental Environment

The hardware configuration used for network training and testing in this article is as follows: 12th Gen Intel(R) Core(TM) i7-12700H (2.3 GHz), 16 GB memory, NVIDIA GeForce RTX 3060, CUDA12.1, Windows 11, Python3.8 and Pytorch2.3. The hyperparameters of the network are: learning rate 0.0001, number of epochs 200, batch size 4, and optimizer Adam.

B. Evaluation of Indicators

Dorsal hand vein classification can be viewed as classifying each pixel in the vein image, with each pixel classification is described by a confusion matrix. The four elements of TP, FN, TN, and FP confusion matrix, where TP is for correctly predicting hand dorsal vein samples as hand dorsal veins, FN for incorrectly predicting hand dorsal vein samples as non-hand dorsal veins, TN for correctly predicting non-hand dorsal vein samples as non-hand dorsal veins, and FP for incorrectly predicting non-hand dorsal vein samples as hand dorsal veins, and the confusion matrix is shown in Table I.

In this paper, the evaluation metrics used for the experimental results are Intersection Over Union (IOU), Mean Intersection Over Union (MIOU), Mean Pixel Accuracy (MPA) and Accuracy for categories. The formulas for IOU, MIOU, MPA and Accuracy are shown in Eq. (9) to Eq. (12), where n denotes the number of categories.

$$IOU = \frac{TP}{TP + FP + FN} \tag{4}$$

$$MIOU = \frac{1}{n}\sum_{i=1}^{n}\frac{TP}{TP + FP + FN} \tag{5}$$

$$MPA = \frac{1}{n}\sum_{i=1}^{n}\frac{TP}{TP + FP} \tag{6}$$

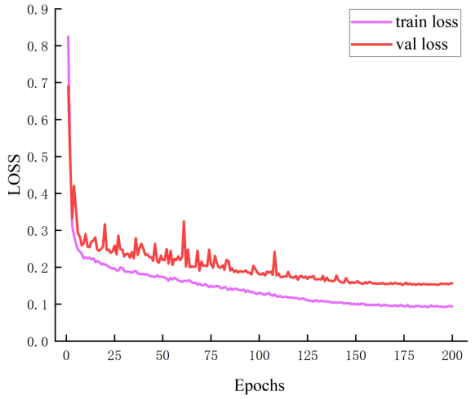
$$Accuacy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

TABLE I. Confusion Matrix of Dorsal Hand Vein Segmentation

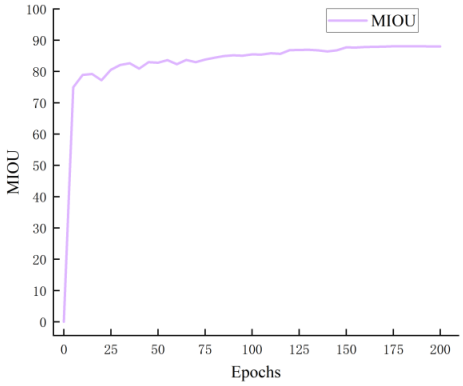
Projected results	Real results	
	hand dorsal vein	Non-dorsal hand veins
hand dorsal vein	TP	FP
Non-dorsal hand veins	FN	TN

C. Training Results of GR-UNet Model

In this paper, the attention mechanism and residual network were introduced to improve U-Net, while a weighted loss function was used to further prevent the imbalance of positive and negative samples. Finally, the improved model was used to train the detection on a self-constructed dataset of hand dorsal veins. The training results of the model are shown in Fig. 12, which indicates that as the number of iterations increased, the accuracy of the model increased while the loss value decreased, and when the loss curve tended to stabilize, the network converged and the training ended. Model accuracy reached a high steady state after 160 iterations.



(a) Loss value.



(b) MIOU.

Fig. 12. Training results of GR-UNet model.

D. Comparative Experiments on Network Models with Different Backbones

In order to verify the impact of backbone feature extraction network on the performance and accuracy of U-Net model, Mobile-UNet, VGG16-UNet and ResNet50-UNet were respectively used for comparative experiments. The U-Net models with different backbone were evaluated by IOU, MIOU, MPA and Accuracy indexes. Results are shown in Table II, where the model performs best when ResNet50 was used as the backbone network.

E. Results of Ablation Experiment

Ablation experiment is a widely used analytical method in machine learning and deep learning to assess the impact of a component of an algorithm, model, or system on overall performance by systematically removing or modifying that component.

The GR-UNet proposed model was experimentally analyzed using a self-constructed dataset of human dorsal hand vein images. An ablation study was conducted utilizing the GR-UNet model to assess the effectiveness of the improved ResNet50 backbone, the incorporation of the GAM attention mechanism at the skip connections, and the enhanced loss function. As demonstrated in Table III, compared to the U-Net model, the IOU of the GR-UNet model increased by 5.28%, the MIoU increased by 2.96%, the MPA increased by 4.14%, and accuracy improved by 0.61%, ultimately reaching 97.5%.

According to the test efficiency statistics of different models presented in Table IV, it is evident that the U-Net model exhibits the shortest processing time; however, its segmentation recognition performance is suboptimal. The implementation of the ResNet50 backbone network, coupled with the integration of the GAM attention mechanism, resulted in increased test time due to the augmented complexity of the network. Nonetheless, this modification significantly enhanced segmentation recognition accuracy. Consequently, the consideration of high-performance computers for model training will be pursued to further improve model efficiency.

F. Comparative Experiments with Different Models

To further validate the segmentation performance of the GR-UNet network model, the U-Net model, ResNet50-UNet model, PSPNet model, and Deeplabv3+ model were employed to assess the recognition capabilities on HDHV dataset. Following image preprocessing, a HDHV image with a consistent size of 288×288 pixels was obtained. The labeled image for training was generated through pixel-by-pixel labeling, with the labeling results presented in a binarized format, where the vein region is indicated in white.

The segmentation recognition results obtained from various network models are illustrated in Fig. 13. The U-Net model was ineffective in extracting weak edges of narrow vessels, leading to instances of local vein fracture. The edge positioning accuracy of the PSPNet model was suboptimal, and its smoothness was inadequate. Although the Deeplabv3+ model outperformed the PSPNet model, it still encountered issues with fracture separation at the connectivity. The ResNet50-UNet addressed the inaccurate localization of edge features seen in the traditional U-Net; however, it continued to exhibit local vein feature breakage. The GR-UNet model, which integrated the ResNet50 network and the GAM attention mechanism, effectively combined the residual structure's performance for feature extraction with the GAM attention mechanism's capability to capture global feature information. Additionally, it improved the loss function to mitigate the imbalance between positive and negative samples, further enhancing the model's segmentation efficacy. Consequently, when compared to other models, the GR-UNet demonstrated superior segmentation performance, excelled in extracting vein vessel boundaries and low-contrast regions, exhibited stronger feature recognition capabilities, and showed good resistance to interference in the segmentation of different HDHV. This further validated the accuracy of the improved network model presented in this paper for dorsal hand vein segmentation.

TABLE II. COMPARISON OF NETWORK MODELS WITH DIFFERENT BACKBONES

Mobile	VGG16	ResNet50	Decoder module	IOU	MIOU	MPA	Accuracy
√	—	—	√	63.31%	79.32%	83.31%	95.69%
—	√	—	√	75.45%	86.09%	92.92%	97.03%
—	—	√	√	77.66%	87.38%	93.35%	97.36%

TABLE III. THE ABLATION EXPERIMENT

Network	ResNet50	GAM	Loss	IOU	MIOU	MPA	Accuracy
Network a	—	—	—	73.54%	85.07%	89.78%	96.89%
Network b	√	—	—	77.66%	87.38%	93.35%	97.36%
Network c	√	√	—	78.81%	88.04%	93.58%	97.52%
Proposed method	√	√	√	78.82%	88.03%	93.92%	97.5%

TABLE IV. TEST EFFICIENCY OF DIFFERENT MODELS

Method	Test time/s	FPS
U-Net	0.0165	60.61
ResNet50-UNet	0.0336	29.76
GR-UNet	0.0579	17.27

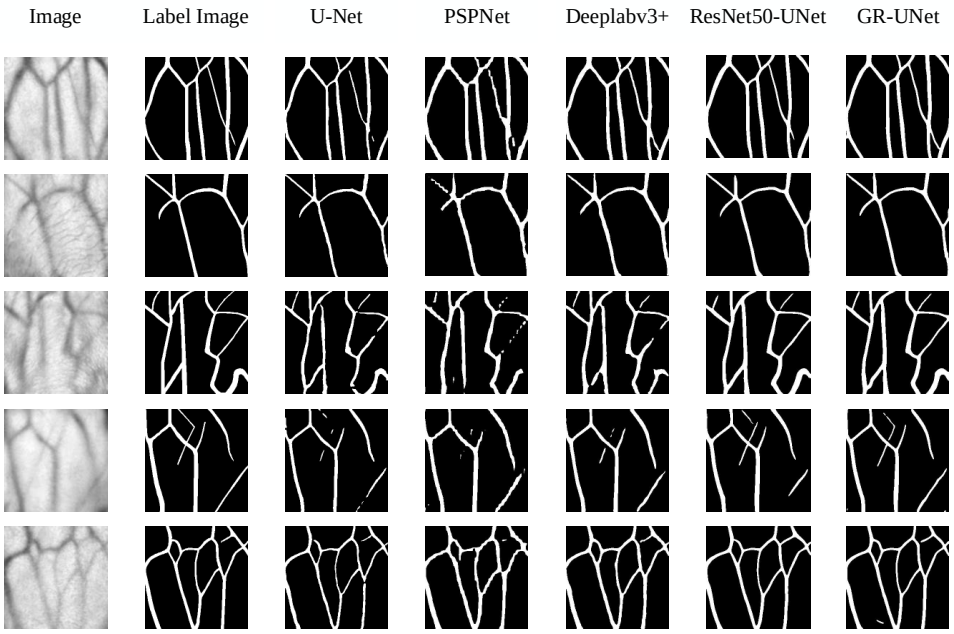


Fig. 13. The results of human back vein segmentation were compared by different models.

In order to verify the performance of the models proposed in this paper, the comparison results of the detection accuracy of each model are shown in Table V. The experimental results showed that the GR-UNet model proposed in this paper outperformed other models in terms of IOU, MIOU, MPA and Accuracy values.

TABLE V. COMPARISON OF DIFFERENT MODELS

Method	IOU	MIOU	MPA	Accuracy
PSPNet	70.29%	83.08%	91.51%	96.24%
Deeplabv3+	73.43%	84.95%	91.82%	96.78%
UNet	73.54%	85.07%	89.78%	96.89%
ResNet50-UNet	77.59%	87.3%	93.52%	97.3%
GR-UNet	78.82%	88.03%	93.92%	97.5%

VI. CONCLUSION

In this paper, a human dorsal hand veins dataset was established by designing a hand dorsal vein visualisation acquisition device based on near-infrared imaging technology. Based on the U-Net model architecture, a GR-UNet model combining attention mechanism and residual network was designed. The model used the residual structure of ResNet50 to enhance the feature extraction capability and effectively solved the vanishing gradient problem. The introduction of the GAM attention mechanism at skip connections optimized the processing of high-level feature information and effectively captured the global contextual features of the network. In addition, a weighted optimization loss function using cross-entropy loss and Dice loss to prevent the positive and negative sample imbalance problem further optimized the performance and stability of the network model.

Compared with the semantic segmentation models of PSPNet, Deeplabv3+, UNet and ResNet50-UNet, the proposed model achieved 78.82% in IOU, 88.03% in MIOU, 93.92% in MPA and 97.5% in Accuracy, respectively. It was superior to other semantic segmentation models, and significantly improved the segmentation accuracy of dorsal hand vein. However, the FPS of the model after completing the training was too low compared to the original model and did not reach the desired value, and a lightweight design of the model will be considered to improve the performance of the model in the subsequent research and there were still some limitations when dealing with more complex backgrounds and low-contrast vein regions, the subsequent work will expand the dataset to improve the generalisation ability and segmentation effect of the model in different age groups, genders and special populations.

ACKNOWLEDGMENT

This work was funded by the Natural Science Project of Zhengzhou Bureau of Science and technology (22ZZRDZX07) and the Scientific Research Special Project of National Clinical Research Base of Traditional Chinese Medicine of Henan Health Commission (2021JDZX2092).

REFERENCES

[1] Saeed A, Chaudhry M R, Khan M U A, Saeed M U, A.Ghfar A, Yasir M N, et al. Simplifying vein detection for intravenous procedures: A comparative assessment through near-infrared imaging system. International Journal of Imaging Systems and Technology, vol. 34, no. 3, p. e23068, 2024.

[2] Ma J, Ye B, Wang S. Vein Image Enhancement Based on CLAHE and Multi-scale Detail Fusion.SEMICONDUCTOR OPTOELECTRONICS, vol. 41, no. 05, pp. 738-742, 2020.

[3] Kuang H, Guan F, Ma X, Liu X. Adaptive threshold vein enhancement method based on illuminance estimation. IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2021, pp. 2591-2595.

[4] Besra B, Mohapatra R K. Extraction of segmented vein patterns using repeated line tracking algorithm. 2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS), 2017, pp. 89-92.

[5] Yakno M, Mohamad-Saleh J, Ibrahim M Z. Dorsal hand vein image enhancement using fusion of CLAHE and fuzzy adaptive gamma. Sensors, vol. 21, no. 19, p. 6445, 2021.

[6] Yang G, Xu X. A FEATURE EXTRACTION METHOD FOR NIR VEIN IMAGE. Computer Applications and Software, vol. 40, no. 04, pp. 199-203+216, 2023.

[7] Zhang H, He L, Wang D. Deep reinforcement learning for real-world quadrupedal locomotion: A comprehensive review. Intelligence & Robotics, vol. 2, no. 3, p. 27597, 2022.

[8] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431-3440.

[9] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer International Publishing, 2015, pp. 234-241.

[10] He T, Guo C, Jiang L, Liu H. Automatic venous segmentation in venipuncture robot using deep learning. 2021 IEEE International Conference on Real-time Computing and Robotics (RCAR), 2021, pp. 614-619.

[11] Lefkovits S, Emerich S, Lefkovits L. Boosting Unsupervised Dorsal Hand Vein Segmentation with U-Net Variants. Mathematics, vol. 10, no. 15, p. 2620, 2022.

[12] Gao X, Zhang G, Zhou F, Yu D. Location Decision of Needle Entry Point Based on Improved Pruning Algorithm. Laser & Optoelectronics Progress, vol. 59, no. 24, pp. 154-166, 2022.

[13] Chen L, Lv M, Cai J, Guo Z, Li Z. U-Net-Embedded Gabor Kernel and Coaxial Correction Methods to Dorsal Hand Vein Image Projection System. Applied Sciences, vol. 13, no. 20, p. 11222, 2023.

[14] Zhao D, Tian Y, Chen H, Zhao Z, Chen Y, Yuan Y. Detection of Dorsal Hand Vein Based on Improved YOLO Nano and Embedded System. Chinese Journal of Biomedical Engineering, vol. 41, no. 06, pp. 691-698, 2022.

[15] Shu Z, Xie Z, Zhang C. Dorsal hand vein recognition based on transmission-type near infrared imaging and deep residual network with attention mechanism. Optical Review, vol. 29, no. 4, pp. 335-342, 2022.

[16] Abd Rahman A B, Juhim F, Chee F P, Bade A, Kadir F. Near infrared illumination optimization for vein detection: hardware and software approaches. Applied Sciences, vol. 12, no. 21, p. 11173, 2022.

[17] Ruan L, Yin Z, Zhou S, Zheng W, Lu W, Zhang T, et al. Vein visualization enhancement by dual-wavelength phase-locked denoising technology. Journal of Innovative Optical Health Sciences, vol. 17, no. 3, p. 2350033, 2024.

[18] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[19] Gao H, Yuan H, Wang Z, Ji S. Pixel transposed convolutional networks. IEEE transactions on pattern analysis and machine intelligence, vol. 42, no. 5, pp. 1218-1227, 2019.

[20] Wu T, Ku T, Zhang H. Research for image caption based on global attention mechanism. Second target recognition and artificial intelligence summit forum, 2020, p. 11427.

[21] Woo S, Park J, Lee J Y, Kweon I S. Cbam: Convolutional block attention module. Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3-19.

A Proposed Batik Automatic Classification System Based on Ensemble Deep Learning and GLCM Feature Extraction Method

Luluk Elvitaria[1], Ezak Fadzrin Ahmad Shaubari[2*], Noor Azah Samsudin[3],
Shamsul Kamal Ahmad Khalid[4], Salamun[5], Zul Indra[6]

Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia[1, 2, 3, 4, 5]
Department of Informatics Engineering-Faculty of Engineering, Universitas Abdurrab, Pekanbaru, Indonesia[1, 5]
Department of Computer Science, Universitas Riau, Pekanbaru, Indonesia[6]

***Abstract*—Classification of batik images is a challenge in the field of digital image processing, considering the complexity of patterns, colors, and textures of various batik motifs. This study proposes an ensemble method that combines texture feature extraction using Gray Level Co-occurrence Matrix (GLCM) with the Residual Neural Network (ResNet) classification model to improve accuracy in batik image classification. Texture features such as contrast, dissimilarity, entropy, homogeneity, mean, and standard deviation are extracted using GLCM and combined with ResNet to produce a more robust classification model. The experimental results show that the proposed method achieves high performance, namely above 90% for each evaluation metric used, such as accuracy, precision, recall and F-1 Score. The best performance in classifying batik images is obtained by the Standard Deviation feature with accuracy, precision, recall, and F1-score of 95%, 93%, 93%, and 93%, respectively. Furthermore, the application of the ensemble method based on the hard voting approach has proven effective in increasing the accuracy of batik image classification by utilizing a combination of texture features and deep learning models. The proposed method makes a significant contribution to the efforts to preserve batik culture through digitalization and can be implemented for various purposes such as an image-based batik search system.**

Keywords*—Batik; GLCM; ResNet; ensemble method; hard voting

I. INTRODUCTION

Batik is an Indonesian cultural heritage that has high artistic and historical value [1], [2], [3], [4]. Each batik motif contains a certain philosophical meaning and can reflect the cultural identity of the producing region. Along with the development of digital technology, digitization and classification of batik images have become important needs, both for cultural preservation and commercialization. However, batik image classification has its own challenges, considering the diversity of complex motifs, colors, and textures [5], [6]. Batik image classification faces challenges in the field of pattern recognition and digital image processing due to the diverse and complex batik motifs [7]. Although there have been various methods applied to classify batik images, such as the use of conventional machine learning-based methods (eg, SVM, KNN) and deep learning (eg, CNN), there are still several limitations that have not been fully overcome.

Many studies only rely on artificial neural networks or deep learning methods for batik image classification without considering specific local texture features, such as those that can be obtained with the GLCM method. This can cause additional problems due to the loss of important information related to patterns and textures that are characteristic of batik motifs. In addition, most previous studies tend to focus on the use of a single technique, either classical feature extraction or deep learning, without combining the advantages of both methods in one ensemble model. This suggests an opportunity to explore a hybrid approach that can improve classification accuracy by leveraging the strengths of GLCM-based texture feature extraction and the generalization capabilities of ResNet.

In this study, it is proposed a batik automatic classification system that combines a texture feature extraction approach using Gray Level Co-occurrence Matrix (GLCM) and an artificial neural network-based classification method, namely Residual Neural Network (ResNet). This ensemble method aims to improve classification accuracy by utilizing the advantages of both approaches. GLCM is known to be effective in describing texture characteristics [8], [9], [10], [11], while ResNet offers advantages in the ability to learn complex feature representations from images through deep network architectures [12], [13]. This research is expected to make a significant contribution to the field of pattern recognition, especially for batik image classification. Furthermore, the presence of this research is expected to provide alternative solutions that are more accurate and efficient in identifying batik images.

By developing an automatic classification system, the digitization and cataloging of Batik patterns can be streamlined, contributing to cultural preservation in the digital age. This also supports educational and archival efforts, making Batik accessible to global audiences. Furthermore, an automated classification system could help streamline various processes, from quality control to market segmentation, and assist in inventory management. For instance, it can aid designers, retailers, and manufacturers in quickly identifying and categorizing different Batik types, improving the overall supply chain and customer service experience.

II. LITERATURE REVIEW

A. GLCM Feature Extraction

Feature extraction is the first and foremost step in image processing, also known as computer vision. This stage involves transforming an image into a set of features that are easier for a computer to understand and relevant for further analysis, such as classification or segmentation [14]. Among the various techniques for feature extraction, Gray-Level Co-Occurrence Matrix (GLCM) which is introduced by Haralick et al. in 1973, is widely used, especially in texture analysis [15], [16]. GLCM is a statistical method that considers the spatial relationship between pixels to analyze texture [17], [18]. GLCM has become a popular choice for feature extraction due to its many advantages. GLCM is able to capture complex texture information in an image by measuring the spatial relationship between pixels [19]. This method allows for more detailed texture analysis compared to simple statistical methods such as mean or variance. It is particularly useful for detecting subtle and complex patterns in images, such as small differences in tissue texture in medical images or variations in surface texture in quality inspection. In addition, GLCM is effective in detecting subtle differences in texture that may not be visible with other analysis techniques [19]. In representing an image, GLCM can produce several features to quantify the texture of an image such as Contrast, Correlation, ASM, homogeneity and so on.

Due to its advantages, GLCM is considered very suitable for image processing in batik because of its ability to analyze textures in depth and capture complex texture patterns. Batik is a traditional fabric that has unique patterns and textures with complex and repetitive motif variations, so it requires effective techniques to recognize, classify, or even detect these patterns. For example, GLCM is able to measure the spatial relationship between pixels in a way that captures the texture properties found in batik images, such as how often a particular pattern appears or how the color intensity changes along the pattern. In addition, GLCM can detect subtle variations in texture with features such as contrast, homogeneity, and energy, which helps identify small patterns and texture differences in batik designs. To date, there have been many studies that applied the GLCM feature extraction algorithm to the task of batik image classification. The summary of the research is shown in Table I.

Based on Table I, it can be seen that GLCM has played a major role in building a batik image classification system. Furthermore, from the literature study that has been conducted, it can be concluded that the majority of these studies apply the k-NN algorithm as their choice. Very few studies have tried to apply Deep Learning-based image classification which is considered one of the best methods in image classification. This is one of the research gaps found and attempted to be solved in this study. A detailed discussion related to this issue will be presented in the following research gap and research contribution sections.

B. Ensemble Method

Ensemble methods have gained significant attention in recent years for the task of image classification due to their ability to improve model accuracy and performance. Ensemble methods are approaches that combine multiple machine learning models to produce a more robust and accurate model [32], [33]. The main goal of ensemble methods is to reduce the prediction error generated by a single model by combining the results of multiple models. Some popular ensemble techniques include Bagging, Boosting, Stacking and Voting [34], [35], [36].

TABLE I. RELATED RESEARCH OF GLCM IMPLEMENTATION FOR BATIK CLASSIFICATION

Authors	Method	Result
[20]	k-NN	Acc = 75%
[21]	k-NN	Acc = 97.96
[22]	Neural Network	Acc = 84%
[23]	SVM	Acc = 78.3%
	k-NN	Acc = 92.3%
[24]	Canberra Distance	Acc = 41.67%
[25]	MLP	Acc = 88.89%
[26]	LVQ	Acc = 98.98%
[27]	Backpropagation Neural Network	Acc = 94%
[28]	Linear Discriminant Analysis	Acc = 96.5%
[29]	Backpropagation Neural Network	Acc = 80%
[30]	k-NN	Acc = 73.33%
[31]	k-NN	Acc = 97%

Bagging (Bootstrap Aggregating) is one of the commonly used ensemble techniques for image classification. This technique works by training several different models on a subset of the training data taken randomly with replacement. Random Forest, one of the Bagging-based algorithms, has been widely used for image classification. Boosting is an ensemble technique that attempts to improve model performance by training the model in stages, where each new model focuses on data that was incorrectly predicted by the previous model. One of the popular Boosting algorithms is AdaBoost (Adaptive Boosting). Stacking is an ensemble technique that combines several learning models to produce a final prediction through a meta-learner model. This technique usually involves heterogeneous models, such as a combination of neural network models and decision tree models. Voting is a relatively simple ensemble technique where several base models are trained independently, and the final prediction is made by voting on the predictions from each model. There are two types of voting-based ensemble methods, Hard Voting and Soft Voting. In Majority Voting (Hard Voting), the final prediction is the prediction that appears most often (the most votes) from all the base models. In contrast to the Weighted Voting (Soft Voting) type, in this type each model is given a weight based on its performance. The final prediction is made based on the weight of each model.

This study chooses the Voting-based ensemble method technique in order to produce a reliable and robust batik image classification system. This is because the voting method has been recognized as a very simple and efficient method [37], [38]. It does not require a lot of complex parameter settings or selection. This makes voting an attractive choice when the main goal is to improve prediction accuracy without adding much complexity. In addition, this Voting method has been proven to be able to overcome the weaknesses of the Single Model which

in many cases has certain weaknesses or biases. By using voting, it can reduce the impact of these weaknesses because the final decision is made based on the consensus of several models. This is very useful when there is uncertainty about which model is most appropriate for a particular dataset. On the other hand, the voting method tends to be more stable and reliable when faced with new or previously unseen data. Because the final result is based on the consensus of several models, this method can adapt better to data that may not follow the training data pattern. However, the implementation of the ensemble method is still rare to produce a batik image classification system. This can be seen in the following Table II.

TABLE II. RELATED RESEARCH OF ENSEMBLE METHOD IMPLEMENTATION FOR BATIK CLASSIFICATION

Authors	Method	Result
[39]	Histogram Feature Extraction with k-NN classifier	Prec = 86.67%
[40]	Ensemble CNN	Acc = 100%
[41]	GLCM + LBP Feature Extration with k-NN classifier	Acc = 93.9%

Based on the study literature summarized in Table II, it can be concluded that the implementation of the ensemble method to produce a robust batik image classification system is still very rare. Especially related to the implementation of the ensemble method combined to GLCM feature extraction and Deep Learning-based classification. Three out of five studies related to the application of the Ensemble Method for batik image classification, it was found that two studies applied the k-NN algorithm as the choice of classification algorithm. Only one study applied a CNN-based Deep Learning algorithm. This problem is another research gap that is tried to be addressed in this study which will be summarized in the following section.

C. Research Gap and Contribution

Regarding the discussion of related research that has been explained in the previous section, it can be concluded that although there have been various methods applied to classify batik images, there are still some limitations that have not been fully resolved. The highlighted meeting problem is the scarcity of research that tries to combine the reliability of GLCM-based feature extraction with Deep Learning classification algorithms. The majority of studies that apply GLCM feature extraction are combined with classical machine learning algorithms such as k-NN instead of utilizing the reliability of the Deep Learning algorithm. The second research gap is the Lack of Ensemble Approaches that Combine Feature Extraction Techniques and Deep Learning. Most previous studies tend to focus on the use of a single technique, be it classical feature extraction or deep learning, without combining the advantages of both methods in one ensemble model. This shows an opportunity to explore a hybrid approach that can improve classification accuracy by utilizing the strengths of GLCM-based texture feature extraction and the generalization capabilities of ResNet. A summary of the research gaps found along with the research contributions proposed by this study is illustrated in Fig. 1.

Based on the illustration related to the research gap illustrated in Fig. 1, it can be concluded that this study proposes two contributions that can bridge those research gap. The first contribution is related to the effort to combine the GLCM feature extraction algorithm which has advantages in recognizing batik textures into a Deep Learning-based classification algorithm. The second contribution attempted by this study is the application of ensemble methods for batik image classification. This study introduces an ensemble method that combines GLCM-based feature extraction techniques and classification using ResNet. By combining these two algorithms, it is hoped that the proposed method can overcome the limitations in capturing complex texture information and fine patterns that are characteristic of batik. Furthermore, through these two contributions, it is hoped that the goal of developing a reliable and robust batik image classification system will be achieved.

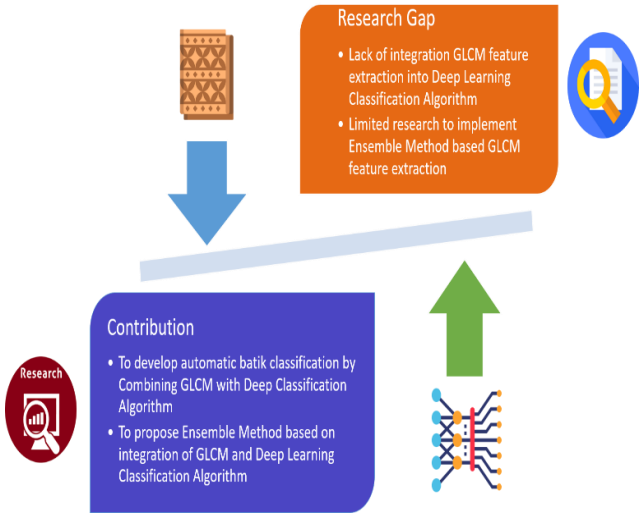


Fig. 1. Research gap and proposed contribution.

III. MATERIAL AND METHOD

A. Dataset Material

The dataset used in this research is a combination of datasets sourced from public datasets such as batik classification Resnet [4] and deep learning batik classification [6]. Furthermore, the dataset used consists of five classes of batik motifs such as lereng, parang, nitik, kawung and ceplok batik, which are illustrated in Fig. 2.

The batik dataset that has been collected consists of 5 classes of 4,284 images and is then separated into three types of data, such as training, testing, and validation data. To assemble data for training, testing, and validation, the dataset was separated by applying a ratio of 70:20:10. The final data for training, testing, and validation can be seen in Table III.

The next step is image preparation from raw data to ready-to-use data, known as the image pre-processing stage. At this image pre-processing stage, two activities are carried out, namely size adjustment and image dataset augmentation. In the image adjustment activity, customization is carried out so that all batik images have a uniform size of 150 x 150 pixels. After that, data augmentation is carried out on the rescaled image using the horizontal rotation method. The overall results of the data collection and pre-processing stages are described in Fig. 3.

Fig. 2. Batik dataset.

TABLE III. DATASET

Batik Class	Data	Training	Testing	Validation
Lereng	405	284	81	41
Parang	1,197	838	239	120
Ceplok	1,053	737	211	105
Kawung	747	523	149	75
Nitik	882	617	176	88
TOTAL	**4,284**	**2,999**	**857**	**428**

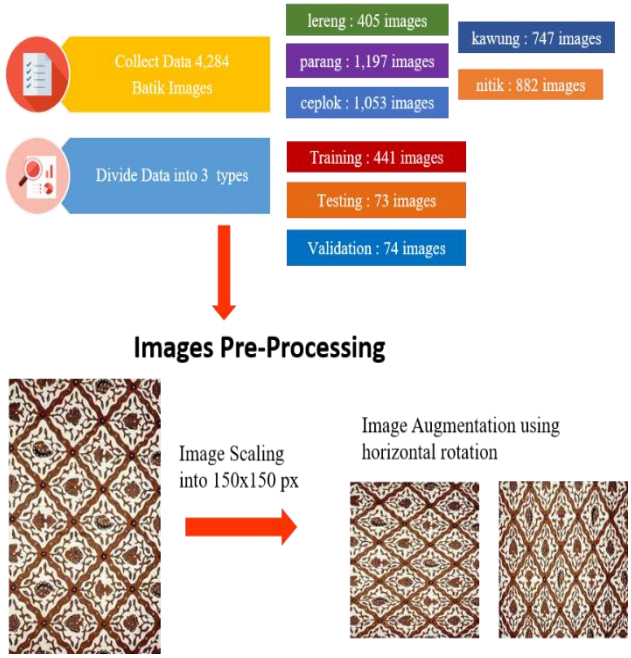


Fig. 3. Image pre-processing stage.

B. Proposed Method

As explained previously, the main contribution of this research is the development of an automatic classification system for batik images by applying the ensemble learning algorithm. Ensemble learning is a method in machine learning that works by combining several machine learning models to produce a more robust and accurate model. The basic concept of ensemble learning is that by combining predictions from several different models, we can reduce errors and improve prediction performance. The concept of ensemble learning applied in this research is based on hard voting. Hard voting is done by calculating the prediction results where the final prediction is determined based on the majority of votes. The overall flow of the proposed ensemble learning for batik image classification is depicted in Fig. 4.

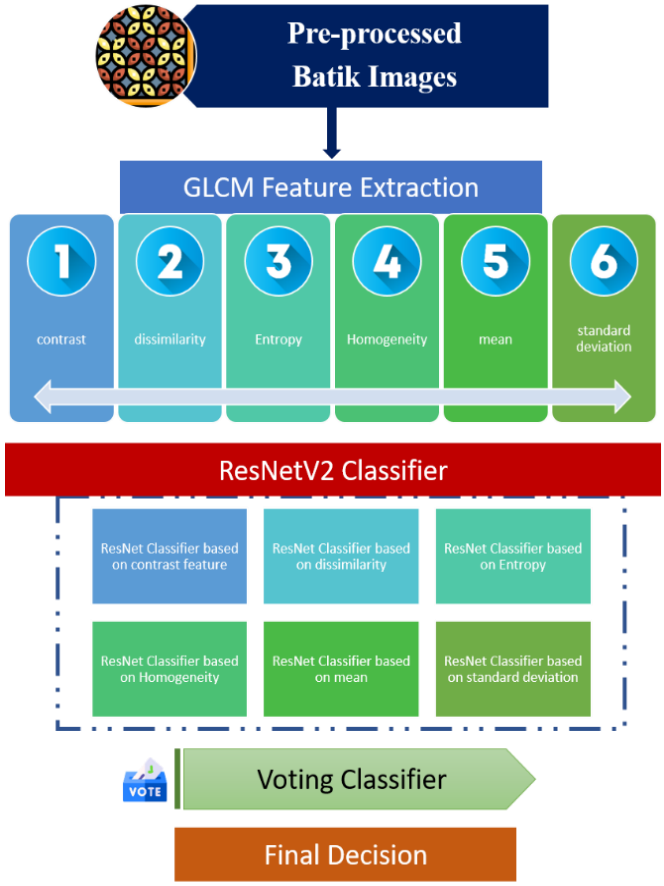


Fig. 4. Flowchart of proposed ensemble deep learning based on GLCM feature extraction.

As seen in Fig. 4, the proposed ensemble deep learning method consists of two main components, namely the GLCM algorithm as a feature extraction layer and the ResNet18 architecture. GLCM (Gray Level Co-occurrence Matrix) is a method often used in image analysis for texture feature extraction. GLCM calculates how often a pair of pixels with a certain intensity value appears in an image at a certain distance and angle. The use of GLCM (Gray Level Co-occurrence Matrix) as a feature extraction algorithm has several advantages that make it a popular choice in image texture analysis. Even when compared to the very popular convolution method, GLCM

has advantages such as the ability to describe texture, various texture features, simple implementation, computational affordability, lower data requirements and so on.

To perform feature extraction using GLCM (Gray Level Co-occurrence Matrix), the first step is to convert the color image to grayscale, because GLCM is applied to grayscale images. After the image is converted, the GLCM parameters are determined, such as the distance between pixels and the angle (0 °, 45 °, 90 °, 135 °). The GLCM matrix is generated by calculating the frequency of occurrence of pixel pairs with certain intensity values at predetermined distances and angles. Once the GLCM matrix is obtained, texture feature extraction is performed by calculating the contrast, dissimilarity, homogeneity, mean, and entropy values of the GLCM matrix. Moreover, this study uses 4 angles at once with a distance between pixels of 1. So that the total features generated for each image are 24 features (4 features for each texture).

After feature extraction with GLCM is completed, the next stage is the application of the CNN algorithm to determine the class of each image by utilizing the Pretrained Residual Network (ResNet) architecture. The ResNet architecture was developed with the aim to overcome the issue of degradation in very deep neural networks. As the depth of the network increases, model performance tends to deteriorate and has difficulty in training due to the vanishing gradient problem. ResNet introduces a residual block that allows information to pass through several layers through "shortcut connections" or "skip connections". ResNet comes in several variants, which differ in the number of layers and network depth. This study applies ResNet18 which is considered suitable for tasks with low data complexity. The block diagram of the ResNet18 architecture is conceived in Fig. 5.

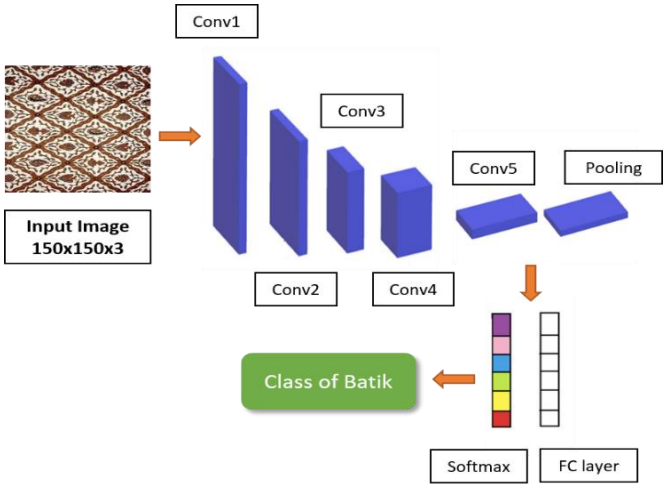


Fig. 5. Block diagram of ResNet18 architecture.

In addition, the ResNet18 architecture is chosen in this study compared to other architectures because of its smaller parameter size. This makes model loading, model weighting, and training much faster. ResNet-18 serves as the base network, and the introduction of the Inception module enhances it. This module combines complex kernels of various sizes to facilitate feature extraction at different image scales, thereby improving recognition accuracy.

The process of identifying the class of batik images will produce six classification results according to the number of textures used. Therefore, an ensemble learning method based on hard voting is then carried out to determine the final class of the batik image. As previously explained, the application of hard voting by combining several different models (ensemble learning) is expected to reduce the risk of overfitting that may occur in a single model. This is because errors made by one model can be offset by other models.

C. Metric Performance

Once the batik image classification model based on ensemble learning was successfully created, the research continued to the performance evaluation stage of the created model by utilizing the confusion matrix. This matrix is very useful for showing the number of correct and incorrect predictions, which are divided into True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). This helps in calculating other metrics and provides deeper insight into the types of errors made by the model. After the Confusion Matrix is successfully identified, the next step is to calculate the level of accuracy, precision, recall and F-1 Score. The illustration of the confusion matrix can be seen in Fig. 6.

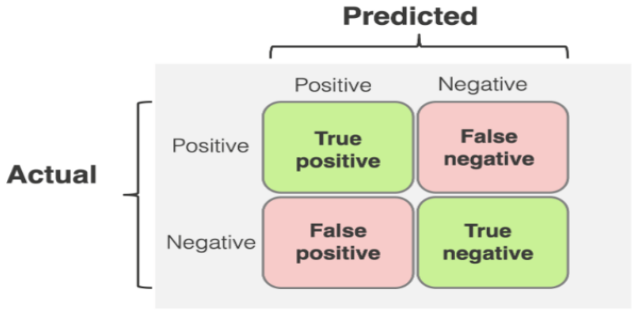


Fig. 6. Confusion matrix scheme.

Each of these metrics has its own advantages and disadvantages. So by using these various metrics, a more comprehensive picture of the ensemble learning model for batik classification will be obtained.

IV. Result and Discussion

As explained, this research applied the concept of ensemble deep learning extraction integrated with CNN algorithm as a feature layer to get the best performance in batik image classification. In this section, a confusion matrix-based performance evaluation will be carried out followed by measuring accuracy, precision, recall and F-1 score. The confusion matrix results for each GLCM feature are shown in Fig. 7 to Fig. 12.

Based on the results depicted by the confusion matrix in Fig. 7 to Fig. 12, it was found that the average GLCM feature has a balanced performance in recognizing batik images. This is indicated by the accuracy of batik class identification between the predicted results and the actual values. Furthermore, there is no significant difference between each GLCM feature used in the classification of batik images. Therefore, it can be concluded that each GLCM feature has a fairly balanced performance in batik classification. Furthermore, by utilizing

the results obtained in the confusion matrix, a performance evaluation based on accuracy, precision, recall and F-1 score was carried out using the following formula.

$$accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\ Score = \frac{2*(Precision*Recall)}{(Precision+Recall)} \quad (4)$$

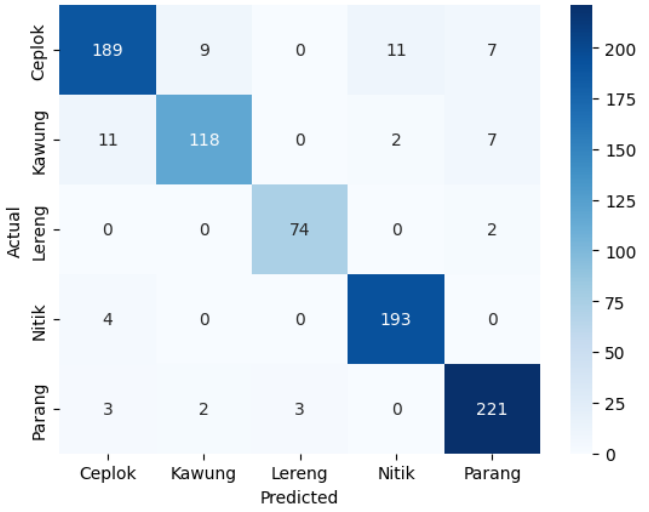


Fig. 7. Confusion matrix based on contrast features.

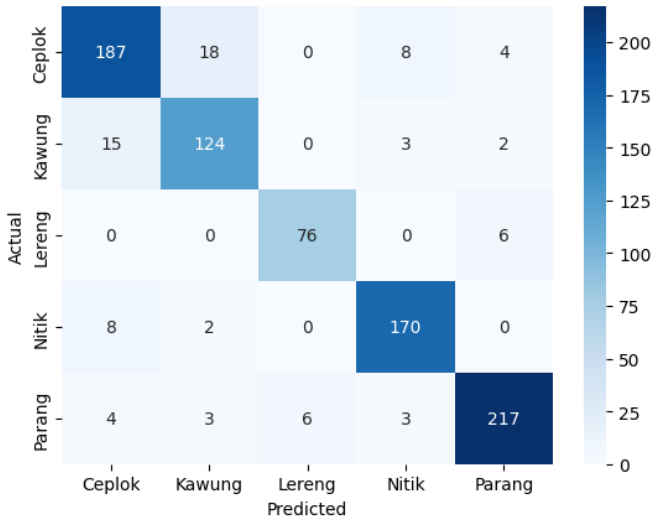


Fig. 8. Confusion matrix based on dissimilarity features.

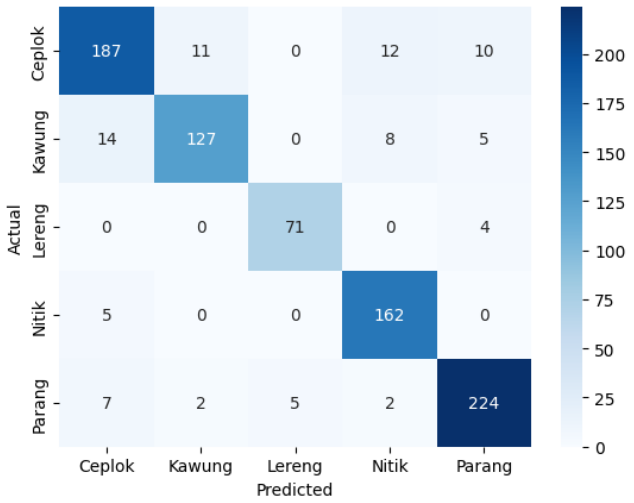


Fig. 9. Confusion matrix based on entropy features.

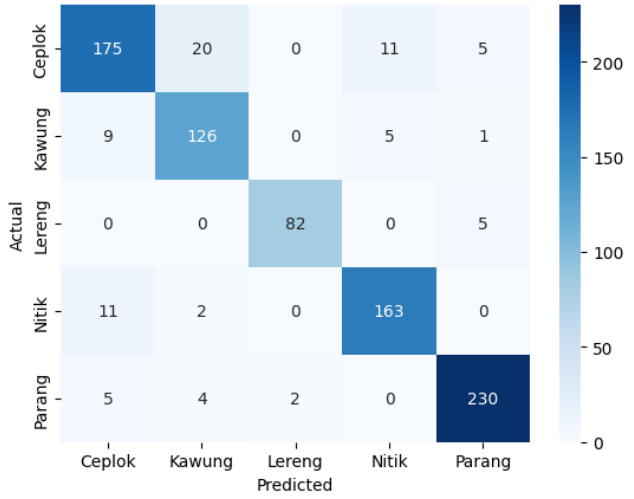


Fig. 10. Confusion matrix based on homogeneity features.

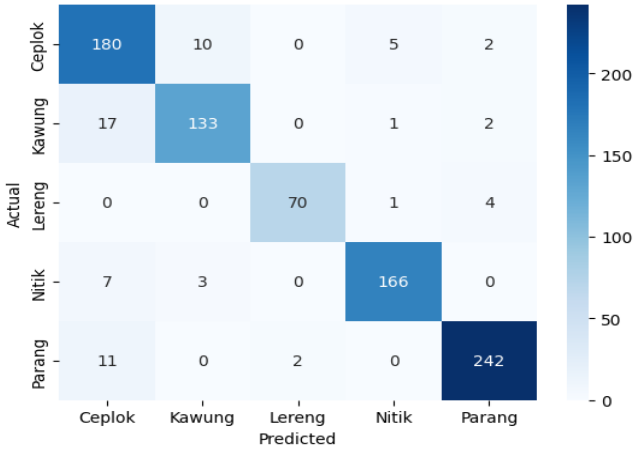


Fig. 11. Confusion matrix based on mean features.

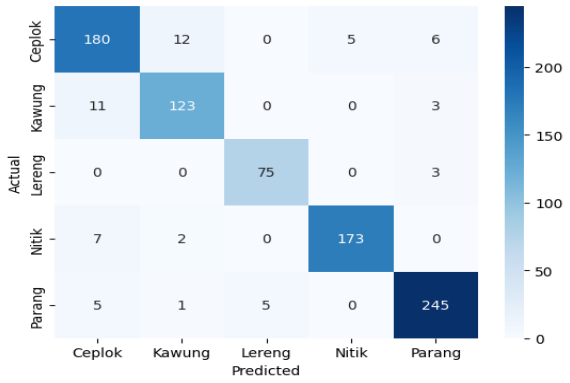


Fig. 12. Confusion matrix based on standard deviation features.

The overall performance evaluation of this research can be seen in Table IV.

TABLE IV. CLASSIFICATION PERFORMANCE FOR EACH GLCM FEATURE

	Acc (%)	Prec (%)	Rec (%)	F1 (%)
Contrast	0.93	0.93	0.90	0.91
Dissimilarity	0.90	0.90	0.90	0.90
Entropy	0.90	0.90	0.90	0.90
Homogeinity	0.95	0.91	0.91	0.90
Mean	0.95	0.92	0.91	0.91
Standard Deviation	0.95	0.93	0.93	0.93

It can be concluded from the performance evaluation results listed in Table II that Standard Deviation is the feature with the best performance because it has the highest value for all metrics (Accuracy, Precision, Recall, and F1-Score) with a value of 93%. Meanwhile, Homogeneity and Mean also show high performance, with the highest accuracy (95%) but have slight differences in precision, recall, and F1-score values. Entropy and Dissimilarity show consistent performance but are lower than other features. In order to improve the classification results, an ensemble method based on hard voting is used. This method is expected to complement each other's shortcomings in the performance of each feature extraction. To test how effective the application of the ensemble method is, a classification of batik images is carried out on the test data. Based on the test, it was found that the application of the ensemble method can improve the performance of the batik image classification system. As an example in a test scenario, by applying the ensemble method based on hard voting, the classification error of the batik class can be avoided. The scenario for testing the effectiveness of the application of the ensemble method is illustrated in Fig. 13.

Based on one of the test results, the prediction results obtained that four out of six features successfully predicted the class of batik images. The other two features failed to identify the class properly. The final result based on majority voting successfully selected the Kawung class which is the right class for the batik image. Therefore, by applying the ensemble method, the performance of the batik classification system can be improved because the final prediction based on the majority decision tends to be more stable and accurate, especially if the data varies. In addition, this ensemble method also has advantages because it can overcome the weaknesses of individual models such as the homogeneity extraction-based classification model which tends to have lower performance in terms of F1 score but has advantages in terms of accuracy. Hence, it can be concluded that by combining predictions from several classification models, the possibility of errors that can occur if only relying on one model can be reduced. This is because the weaknesses of one classification model can be compensated by another model. Although the study has successfully proven that the implementation of the ensemble method based on the GLCM feature extraction algorithm and the ResNet architecture is able to produce a robust system for producing a batik image classification system, there are many more things that need to be improved in further research. Recommendations for further research will be discussed in the following conclusion section.

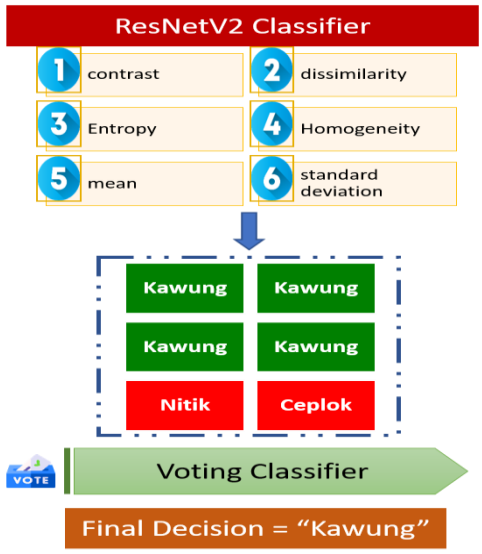


Fig. 13. Testing scenario for ensemble methods implementation.

V. CONCLUSION

The study exploited image extraction algorithms to obtain texture images that can be used in batik image classification tasks. The image extraction algorithm used is GLCM which is known to be reliable in describing the texture of images effectively. Furthermore, the application of GLCM produces 6 features including Contrast, Dissimilarity, Entropy, Homogeneity, Mean, and Standard Deviation. Based on the results of applying these features to classify batik images, several interesting findings were obtained. The Standard Deviation feature showed the best performance with accuracy, precision, recall, and F1-score of 95%, 93%, 93%, and 93%, respectively. This shows that this feature is very good at capturing important characteristics of batik motifs. The Mean and Homogeneity features also showed very good performance, with an accuracy of 95% and other metric values slightly lower than the Standard Deviation feature. On the other hand, the Contrast feature showed quite high accuracy (93%), but had a lower recall value (90%), indicating good performance but not as good as the top three features. The last finding related to the Dissimilarity and Entropy Features, these two features were

recorded to have similar performance with accuracy, precision, recall, and F1-score values of 90%, indicating that these features are less effective than other features.

In addition, to improve the performance of batik image classification, the application of the Ensemble method based on majority voting (hard voting) is proposed. By combining the results of several models trained using different features through the Majority Voting method, the performance of batik image classification can be improved. This method utilizes the advantages of each feature, thereby increasing the accuracy and stability of the model as a whole. The Ensemble method with Majority Voting shows that although some individual models may have lower performance, combining predictions through majority voting produces a more reliable and accurate model. By utilizing the diversity of information captured by various texture features, the ensemble method helps reduce prediction errors and improves the generalization ability of the model.

This study shows that the use of ensemble method with Majority Voting is an effective approach for batik image classification. The combination of texture features Standard Deviation, Mean, and Homogeneity gives the most optimal results, but the ensemble of various features as a whole increases the accuracy and reliability of the model. These results emphasize the importance of using ensemble methods to handle complex image classification problems and support practical applications such as automatic motif recognition for batik products. As a continuation of the research, it is interesting to conduct experiments by applying various other ensemble methods such as bagging, boosting to find the ensemble method with the best performance. In addition, from the classification algorithm side, the application of other algorithms such as Random Forest, Xgboost, SVM and so on. Hence, it can be obtained automatic classification system which is truly reliable and robust in recognizing batik images.

REFERENCES

[1] W. Steelyana, "Batik, a beautiful cultural heritage that preserve culture and support economic development in Indonesia," Binus Bus. Rev., vol. 3, no. 1, p. 116, 2012.

[2] N. Akagawa, "Batik as a creative industry: Political, social and economic use of intangible heritage," in Safeguarding Intangible Heritage, Routledge, 2018, pp. 135–154.

[3] T. Sekimoto, Batik as a Commodity and a Cultural Object, vol. 1. Berghahn Books, 2003.

[4] C. Wang, "Building a network for preserving intangible cultural heritage through education: A study of Indonesian batik," Int. J. Art Des. Educ., vol. 38, no. 2, pp. 398–415, 2019.

[5] I. Nurhaida, A. Noviyanto, R. Manurung, and A. M. Arymurthy, "Automatic Indonesian's batik pattern recognition using SIFT approach," Procedia Comput. Sci., vol. 59, pp. 567–576, 2015.

[6] E. Winarno, A. Septiarini, W. Hadikurniawati, and H. Hamdani, "The Hybrid Features and Supervised Learning for Batik Pattern Classification," ACM J. Comput. Cult. Herit., vol. 17, no. 2, pp. 1–13, 2024.

[7] N. D. Girsang, "Classification Of Batik Images Using Multilayer Perceptron With Histogram Of Oriented Gradient Feature Extraction," in Proceeding International Conference on Science and Engineering, 2021, pp. 197–204.

[8] P. Mohanaiah, P. Sathyanarayana, and L. GuruKumar, "Image texture feature extraction using GLCM approach," Int. J. Sci. Res. Publ., vol. 3, no. 5, pp. 1–5.

[9] A. H. Farhan and M. Y. Kamil, "Texture Analysis of Breast Cancer via LBP, HOG, and GLCM techniques," in IOP conference series: materials science and engineering, IOP Publishing, 2020, p. 72098.

[10] M. Rafi and S. Mukhopadhyay, "Texture description using multi-scale morphological GLCM," Multimed. Tools Appl., vol. 77, pp. 30505–30532, 2018.

[11] Z. Indra and Y. Jusman, "Performance of GLCM algorithm for extracting features to differentiate normal and abnormal brain images," in IOP Conference Series: Materials Science and Engineering, IOP Publishing, 2021, p. 12011.

[12] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," Artif. Intell. Rev., vol. 53, pp. 5455–5516, 2020.

[13] M. Shafiq and Z. Gu, "Deep residual learning for image recognition: A survey," Appl. Sci., vol. 12, no. 18, p. 8972, 2022.

[14] D. Weinland, R. Ronfard, and E. Boyer, "A survey of vision-based methods for action representation, segmentation and recognition," Comput. Vis. image Underst., vol. 115, no. 2, pp. 224–241, 2011.

[15] G. Prasad, G. S. Vijay, and R. Kamath, "Comparative study on classification of machined surfaces using ML techniques applied to GLCM based image features," Mater. Today Proc., vol. 62, pp. 1440–1445, 2022.

[16] G. Beliakov, S. James, and L. Troiano, "Texture recognition by using GLCM and various aggregation functions," in 2008 IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence), IEEE, 2008, pp. 1472–1476.

[17] Ş. Öztürk and B. Akdemir, "Application of feature extraction and classification methods for histopathological image using GLCM, LBP, LBGLCM, GLRLM and SFTA," Procedia Comput. Sci., vol. 132, pp. 40–46, 2018.

[18] S. Barburiceanu, R. Terebes, and S. Meza, "3D texture feature extraction and classification using GLCM and LBP-based descriptors," Appl. Sci., vol. 11, no. 5, p. 2332, 2021.

[19] M. Hall-Beyer, "Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales," Int. J. Remote Sens., vol. 38, no. 5, pp. 1312–1338, 2017.

[20] D. Wijaya and A. R. Widiarti, "Batik classification using KNN algorithm and GLCM features extraction," in E3S Web of Conferences, EDP Sciences, 2024, p. 2012.

[21] R. Andrian, M. A. Naufal, B. Hermanto, A. Junaidi, and F. R. Lumbanraja, "K-Nearest Neighbor (k-NN) Classification for Recognition of the Batik Lampung Motifs," in Journal of Physics: Conference Series, IOP Publishing, 2019, p. 12061.

[22] K. Chandraprabha and S. Akila, "Texture feature extraction for batik images using glcm and glrlm with neural network classification," Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol, vol. 5, pp. 6–15, 2019.

[23] A. E. Minarno, F. D. S. Sumadi, H. Wibowo, and Y. Munarko, "Classification of batik patterns using K-Nearest neighbor and support vector machine," Bull. Electr. Eng. Informatics, vol. 9, no. 3, pp. 1260–1267, 2020.

[24] Y. Sari, M. Alkaff, and R. A. Pramunendar, "Classification of coastal and Inland Batik using GLCM and Canberra Distance," in AIP Conference Proceedings, AIP Publishing, 2018.

[25] N. K. Hamzidah, A. Jariyah, A. R. Ramadhani, N. Nurhasni, M. M. Parenreng, and S. Suyuti, "Evaluation of image feature extraction using gray level co-occurrence matrix (GLCM) parameters and multilayer perceptron (MLP) algorithms in classifying typical batik motifs of South Sulawesi," in AIP Conference Proceedings, AIP Publishing, 2024.

[26] E. M. Yuniarno and M. H. Purnomo, "Indonesian batik image classification using statistical texture feature extraction Gray Level Co-occurrence Matrix (GLCM) and Learning Vector Quantization (LVQ)," J. Telecommun. Electron. Comput. Eng., vol. 10, no. 2–3, pp. 67–71, 2018.

[27] C. S. K. Aditya, M. Hani'Ah, R. R. Bintana, and N. Suciati, "Batik classification using neural network with gray level co-occurence matrix and statistical color feature extraction," in 2015 International Conference on Information & Communication Technology and Systems (ICTS), IEEE, 2015, pp. 163–168.

[28] A. E. Minarno, I. Soesanti, and H. A. Nugroho, "Batik Classification using Microstructure Co-occurrence Histogram," JOIV Int. J. Informatics Vis., vol. 8, no. 1, pp. 134–140, 2024.

[29] A. R. Juwita and A. Solichin, "Batik pattern identification using GLCM and artificial neural network backpropagation," in 2018 Third International Conference on Informatics and Computing (ICIC), IEEE, 2018, pp. 1–6.

[30] C. Irawan, E. N. Ardyastiti, E. H. Rachmawanto, and C. A. Sari, "A survey: Effect of the number of GLCM features on classification accuracy of lasem batik images using K-nearest neighbor," in 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), IEEE, 2018, pp. 33–38.

[31] F. U. Karimah and A. Harjoko, "Classification of batik kain besurek image using speed up robust features (SURF) and gray level co-occurrence matrix (GLCM)," in Soft Computing in Data Science: Third International Conference, SCDS 2017, Yogyakarta, Indonesia, November 27–28, 2017, Proceedings 3, Springer, 2017, pp. 81–91.

[32] M. RE and G. VALENTINI, "Ensemble methods in data mining: improving accuracy through combining predictions," Comb. pattern Classif. Wiley, Hoboken, pp. 186–229, 2012, doi: 10.1201/b11822-34.

[33] M. Zounemat-Kermani, O. Batelaan, M. Fadaee, and R. Hinkelmann, "Ensemble machine learning paradigms in hydrology: A review," J. Hydrol., vol. 598, p. 126266, 2021.

[34] L. Wen and M. Hughes, "Coastal wetland mapping using ensemble learning algorithms: A comparative study of bagging, boosting and stacking techniques," Remote Sens., vol. 12, no. 10, p. 1683, 2020.

[35] I. D. Mienye and Y. Sun, "A survey of ensemble learning: Concepts, algorithms, applications, and prospects," IEEE Access, vol. 10, pp. 99129–99149, 2022.

[36] A. Jurek, Y. Bi, S. Wu, and C. Nugent, "A survey of commonly used ensemble-based classification techniques," Knowl. Eng. Rev., vol. 29, no. 5, pp. 551–581, 2014.

[37] K. T. Leung and D. S. Parker, "Empirical comparisons of various voting methods in bagging," in Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 2003, pp. 595–600.

[38] S. Saha and A. Ekbal, "Combining multiple classifiers using vote based classifier ensemble technique for named entity recognition," Data Knowl. Eng., vol. 85, pp. 15–39, 2013.

[39] H. Ramadhan, I. Arieshanti, A. Yuniarti, and N. Suciati, "Impression Determination of Batik Image Cloth By Multilabel Ensemble Classification Using Color Difference Histogram Feature Extraction," J. Ilm. Kursor, vol. 7, no. 4, 2014.

[40] Y. Azhar, M. C. Mustaqim, and A. E. Minarno, "Ensemble convolutional neural network for robust batik classification," in IOP Conference Series: Materials Science and Engineering, IOP Publishing, 2021, p. 12053.

[41] F. T. Kurniati, D. H. F. Manongga, I. Sembiring, S. Wijono, and R. R. Huizen, "The object detection model uses combined extraction with KNN and RF classification," arXiv Prepr. arXiv2405.05551, 2024.

A Comprehensive Crucial Review of Re-Purposing DNN-Based Systems: Significance, Challenges, and Future Directions

Yaser M Al-Hamzi, Shamsul Bin Sahibuddin
Faculty of Artificial Intelligence (FAI), University Technology Malaysia (UTM), UTMKL, Kuala Lumpur, Malaysia

***Abstract*—The fourth industrial revolution is marked by the significance of artificial intelligence (AI), particularly the remarkable progress in deep neural networks (DNNs). These networks have become crucial in various areas of daily life because of their remarkable pattern-learning capabilities on massive datasets. However, the incompatibility of these systems makes reutilizing them for efficient data analysis and computation highly intricate and challenging due to their fragmentation, internal structure, and complexity. Training in DNNs, a vital essential activity in model development, is often time-consuming and costly intensive computation. More precisely, reusing the entire model during deployment when only a small portion of its required features will result in excessive overhead. On the other hand, reengineering the model without efficient code review could also pose security risks as the model would inherit its defects and weaknesses. This paper comprehensively reviews DNN-based systems, encompassing cutting-edge frameworks, algorithms, and models for complex data and existent limitations. The study, which results from a thorough examination, analysis, and synthesis of observations from 193 recent scholarly papers, provides a wealth of knowledge on the subject, identifying key issues and future research directions by offering novel guidelines to advance the DNN model's repurposing and adaptation, especially in finance, healthcare, and autonomous applications. The demonstrated findings, specifically those related to failure and risk challenges of DNN converters, including factors (n=12), symptoms (n1=4, n2=3), and root causes (n1=4, n2=3), will enrich the ML-DNNs community and guide them toward desirable model development and deployment improvement, with significant practical implications for intelligent industries.**

***Keywords*—DNNs; DNN-based systems; significance and challenges; incompatibility; re-purposing; review**

I. INTRODUCTION

In the digital age, computer science has increasingly focused on the advancements of artificial intelligence, which has become a thriving area of research [1]. However, despite the urgency of matching the rapid development of the fourth industrial revolution, literature overlooks the scarcity of resources related to AI, ML, and DNN-based systems [2] , [3]. It fails to discuss their importance within ecosystems and their various difficulties [4], [5], [6]. The ability to analyze and forecast potential growth through complex data analytics, leveraging deep learning capabilities, leads to increased business value [7]. More specifically, convolutional neural networks CNN [8] and recurrent neural networks RNN [9], the two primary types of DNN architectures, have been intensively investigated to handle various NLP [10] and computer vision CV [11] Problems. It has been demonstrated that different hyperparameters (e.g. learning rate, number of layers, epochs number, optimizer, hidden size, batch size, and regularization techniques) can cause DNN performance to vary significantly [12]. In addition, optimizing parameters and ensuring compatibility between CNNs and RNNs is critical for their performance [13]. This is because these techniques are relatively new and constantly evolving, with ongoing research exploring how best to leverage their combined strengths [12], [13]. However, while several engineering efforts attempted to overcome the exchange complexities between artificial architectures, specifically machine learning (ML) and deep learning (DL) techniques, the literature indicates that significant limitations still exist, whether in domain knowledge or practical applications [1], [14], [15], [16], [17]. For instance, the inspiration of technology reuse [18], [19], [20], [21], and [22] demonstrated that one of the most significant challenges facing (ML) and (DL) developers, researchers, and end users is the lack of interoperability between their systems [13].

In more detail, reusing DNN-based systems computation is difficult, as with any emerging technology [23], [3]. One of these obstacles is the structural problem of non-interoperability among DNN-based systems, along with a need for more technical skills and engineering methods. [3]. Therefore, this systematic review examines cutting-edge interfaces, models, frameworks, and algorithms tightly associated with machine learning (ML) and deep learning (DL) approaches [24].

Importantly, this study stands out as the first systematic investigation that comprehensively addresses the existing open issues in machine and deep learning systems and related technologies to the best of our knowledge. It explores prominent techniques like "Facebook's Torch-PyTorch and Caffe2" [16], [25], [7], [26], [21], [20], "Montreal University's Theano, TensorFlow founded by Google [27]; [28], Apache's MxNet, Microsoft's CNTK [21], [20], [28], [29], and Hugging-Face [30]" highlighting their contributions to the analysis diverse data types, including complex data as well as the existing limitations. The systematic investigation included a pack of widely used deep neural network models like LeNet-1 [31], LeNet-5 [32], ResNet-18 and ResNet-152 [33], Xception [34], Inception-V3 [35], VGG-16 [36], VGG-19 [37]. The findings of this review illustrated that, in development and deployment processes, the lack of compatibility has three categories: software level, hardware level, and architecture level. The software level includes "type of used programming

language" [4], [38], type of "ML and DL framework" [24], type of "ML and DL model" [39], type of" ML and DL algorithm" [12], and type dataset [40], [41]). The hardware level encompasses the type of "computer manufacturer, processor, and type of accelerators (e.g., GPUs, TPUs, FPGAs) [42], [43]. The architecture level comprises structure design and mission misalignment levels [44], [45], [46]. In a relevant context, the findings of this review indicate that a unified model that combines two or more of these methods can significantly enhance deep neural network performance (e.g., memory consumption and inference time) [24], [44].

This review explores the potential of (DNNs) and the challenges they present in the reuse context, enabling more efficient progress in methodologies and processes. It comprehensively addresses their significance and reuse challenges from a unique perspective. The aim is to make a crucial contribution to advancing DNN-based systems. The promising unified approach with its novel techniques can effectively promote compatibility and reuse process, leading to obtaining the desirable accuracy [47], [48], [49] and robustness [50], [49] towards various types of adversarial attacks, reducing computing time and lower computing costs [48], [51], [24].

II. Deep Neural Networks Reuse

A. Overview

The reuse approach involves utilizing existing software engineering and artificial intelligence technologies for different purposes, such as reducing engineering and computation costs and time [23], [52]. Nevertheless, reusability from a deep neural network perspective has received limited attention [53], [54]. One of the main focuses of this paper is the incompatibility problem, challenges, and possible solutions for reusing deep neural networks (DNNs). We do this by highlighting four main reuse paradigms that have been identified through literature review, analysis, and synthesis. These paradigms consider things like the need for computing resources and inference time, hardware configurations, and dataset dependencies [23], [55], [3]. The four include conceptual reuse, development and assessment of the need for reuse, adaptation to reuse, and deployment reuse, as illustrated in Fig. 1.

B. Reuse Paradigms Definition

1) Conceptual reuse: It involves replicating and reengineering algorithms or model architectures from academic literature, often due to licensing or using a specific DL framework. This approach is related to Sommerville's abstraction reuse. [56] and is crucial for scientific reproducibility [57].

2) Model development and reusability assessment: Involve accurately determining the nature of the intended task, followed by data preprocessing, hardware preparation, and algorithm selection for training a model from scratch and evaluating the need and potential for reuse. [58]; [30].

3) Adaptation reuse: Utilizing existing DNN models for different learning tasks, leveraging techniques like transfer learning or knowledge distillation. [59]. This approach is suitable for publicly available pre-trained models PTMs, allowing engineers to customize them for different tasks. [60].

4) Deployment reuse: This method of reusing pre-trained DNN models in various computational environments and frameworks is ideally suited for an engineer's desired task. This approach is similar to Sommerville's "system reuse" and involves fine-tuning, then converting the model from one representation to another, followed by compilation to optimize for hardware. Multiple forms of reuse can be possible in a single engineering project [3].

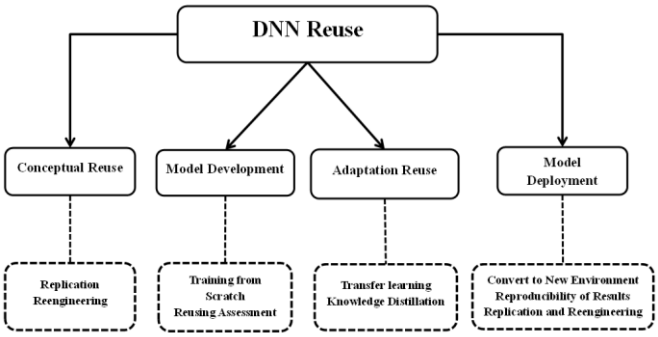


Fig. 1. The four paradigms of deep neural network reuse. The information illustrated in dashed boxes is an example of each paradigm type.

III. Motivation

Machine learning and deep neural networks have benefited numerous aspects of AI, including finance, autonomous applications, and healthcare (e.g. drug discovery, disease predictions, and medical image analysis in intelligent healthcare applications) [61]. However, a critical barrier prevents AI from reaching its full potential. The issue concerns the incompatibility of AI environments, mainly when leveraging previous successful efforts like drawing inspiration from developed codes and pre-trained models [17]; [13]. This incompatibility partially splits the development landscape and prevents code and model sharing, which is an effective way to boost innovations within the field [62]. Reusing existing code and pre-trained models provides significant benefits as follows:

A. Reduced Development Time

For instance, [23] and [54], demonstrated that, by reusing existing code and pre-trained models, experts could focus on possible and innovative solutions, thereby saving significant time that would otherwise be required to construct all components from scratch.in addition, reutilizing eliminates the need to compute similar data multiple times, speeds up development, and reduces costs.

B. Enhanced Reproducibility

Code generation, model sharing, and reusing approaches improve the overall reproducibility of research, allowing for verifying and enhancing existing solutions [13], [55].

C. Accelerated Innovation

By extension, extending previous frameworks and models means that researchers can work on more complicated solutions over a shorter period and advance the field of AI [52]. However, current approaches to reusing and reengineering the previous

ML and DNN productions encounter certain limitations, such as:

1) Performance bottlenecks: Reused models may consume more memory during inference, which may take longer than inference in their native environment [23].

2) Concerns over accuracy and robustness: Fine-tuning necessitates a series of decisions to maintain the model's accuracy in the newly applied context [49].

3) Converting the entire model challenge: Existing methods are ineffective because they require reloading the whole model rather than just the necessary portions, resulting in costly and time-consuming computational resources [23]; [63].

4) Inheritance model defects, when reused, may cause security vulnerabilities: When reusing models, the new system may inherit threats from the previous models, thereby increasing its vulnerability to attacks. These limitations thus advocate for extensive research and development of code and model reuse.

To overcome these challenges, this study suggested the following critical solutions:

- Standardized Frameworks: Many other research areas are challenged to create coherent architectures that facilitate simple integration of off-the-shelf models and built-from-scratch submodules.
- Advanced Transfer Learning Techniques: Studies on techniques for implementing transfer learning, such as domain adaptation, fine-tuning methodologies, and lifelong learning paradigms.
- Developing Novel, Unique Reuse Methods: Involves exploring strategies for reusing only specific aspects of a model, which enables the creation of unique solutions without requiring additional work.
- Development of Security-Aware Reuse Strategies: Creating strategies for distinct discovery and protection of security threats whenever a model is reused to create other new systems. Solving these issues will lead to a healthier ground for code and model sharing in AI. In the long run, this will lead to more innovation, improve the use of AI to its full potential, and bring about changes that will positively impact society.

IV. Scope of Study and Design Architecture

In general, the search strategy for addressing the main research problem and the related challenges in this systematic review is to move in an inverted pyramid, i.e. from the broadest level to the narrowest one, as shown in Fig. 2. The investigation and synthesis then continue to narrow coherently, ultimately reaching the most specific, valid, and critical missing points that previous studies and current methods and approaches have neglected. We examine, investigate, synthesize, analyze, and discuss the most significant open issues and challenges that are closely related to the main research focus of incompatibility, drawing inspiration from the reuse approach and the limitations of current systems reengineering methods. We initially go from the big picture of AI, focusing on computer vision CV and natural language processing NLP, to the more specific view of machine learning with its related techniques (e.g. algorithms, models, and frameworks). Then we move to the much more specific domain of deep neural networks (DNNs), which includes the most computational techniques, such as CNN-based, RNN-based, and hybrid systems, such as DNN-NLP, DNN-CV, NLP-CV, and DNN-NLP-CV [64], [65], [66].

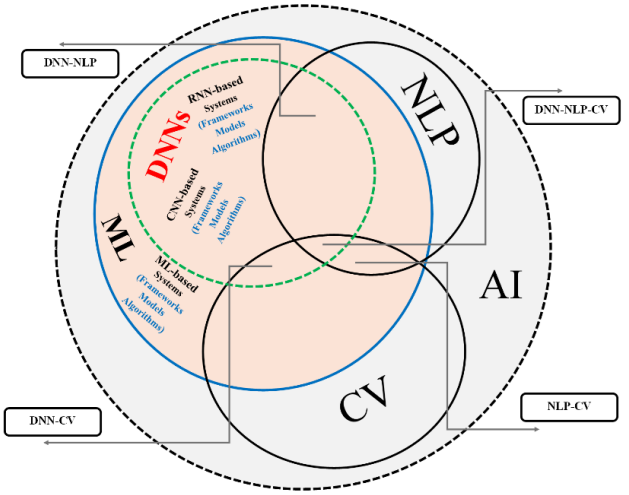


Fig. 2. The architecture of the dependence between AI, ML, computer vision (CV), NLP, and Hybrid DNN-based systems.

V. Concepts

A. The Border Domain of Artificial Intelligence (AI)

Artificial Intelligence (AI) is a "branch of computer science," and its latest generation is rapidly expanding. This makes it an attractive topic for research that focuses on designing intelligent machines that mimic human thought processes, such as learning and decision-making [1]. This level of automation means that such machines give performances on inference, categorization, and detection for activities formerly involving interfaces [67]. The AI context includes the facilities and technologies used to build intelligent applications and release them into the environment [68].

B. Machine Learning (ML)

ML can be defined as a category of AI involving the implementation of algorithms and methods that enable the computer to learn from data and experience without presenting a set of instructions [69]. DNN-based models can learn more about a particular problem as opposed to other approaches by mirroring the data and identifying patterns and relationships that can be used to construct models [70]. The most frequent categories of ML problems entail classification, regression, and clustering. According to [67], although machine learning models have brought significant improvements to AI applications in fields ranging from finance and healthcare to manufacturing, this progression was predominantly due to their relative accuracy (concerning traditional linear statistics classifiers) when it comes to making predictions. However, this benefit comes with a drawback: the lack of transparency in the design of machine learning models often leads to their characterization as "black boxes". On the other hand [71], demonstrated that the integration of (ML) and (AI) components

into public sector applications faces significant limitations due to their fragility and algorithmic mismatches.

C. Deep Neural Networks (DNNs)

Deep learning, also known as "deep neural networks DNN," is a more advanced form of machine learning that requires using an artificial neural network with multiple layers embedded [72]. These deep networks, which optimize data-intensive tasks like image, text, voice, and speech recognition, are based on the structure and functioning of the human brain [73], [74], [16]. With deep learning algorithms, changes are made to the connections between artificial neurons in the network, allowing them to capture progressively more complex features in the data. [75].

Moreover, Recently, DNNs have demonstrated remarkable achievements in the medical field, such as the diagnosis and prediction of diseases such as Alzheimer's disease, heart disease, lung and liver cancer, kidney and brain cancer, and many more. In this context, we will focus on the role of DNN-based algorithms in the diagnosis and prediction of Alzheimer's disease [76], [77], [78]. Furthermore, it has demonstrated remarkable results for low-data drug detection despite the limitation of "out-of-domain generalization" [79]. However, the challenge of incompatibility among ANN-based technologies and the real world is addressed in this study [80].

The fragmentation of models, algorithms, tools, libraries, and frameworks leads to code and model reusability issues. [48], [4], [81]. We also highlight the lack of systematic approaches to enhance interoperability, reduce complexity, and explain mechanisms in artificial neural networks. [82]. Fig. 3 illustrates the basic architecture of the perceptron, as shown in part (a), in DNNs with three input neurons, four neurons in each hidden layer, and two neurons in output layers, as shown in parts (b) and (c) [64]; [65]; [66].

D. The Neuron

Equation 1 describes the calculation of the neural network's output value, or activation, which involves summarizing the activations of all neurons from the "previous layer" connected to the evaluated neuron, adding the neuron's bias, and applying the "activation function" to produce the final neuron activation.

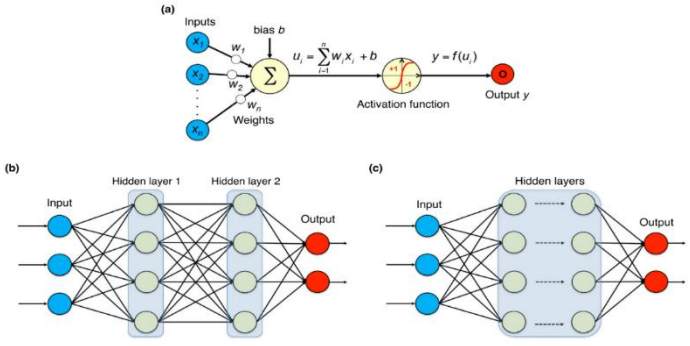


Fig. 3. The basic architecture of Multi-Layer "Perceptron" part (a) and the (DNNs) input and output layers, as shown in parts (b) and (c).

This process introduces non-linearity between inputs and outputs, ensuring the network's outputs are linear combinations of inputs. The process of adjusting weights and biases to achieve desired results is called training, where the weights and biases of the input neurons determine the network's output.

$$U_i = \sigma\left(\left(\sum_j^n U_j * w_{ij}\right) + b_i\right) \quad (1)$$

Where U_i : evaluated neuron activation, σ: activation-function, n: Set of input neurons linked to evaluated neuron, U_j: activation of input neurons from the previous layer, w_{ij}: weight of the connection between neurons, b_i: Bias value of evaluated neuron.

E. Transfer Learning

The transfer learning approach is a popular and effective DL technique that uses pre-trained models PTMs to address challenging issues [83]. However, choosing the best-trained model for a target job remains difficult as most properly training every candidate model is a high computationally cost and long inference time, highlighting the critical necessity for a convenient prediction metric based solely on early training outcomes [81], [3].

F. Model Development and Reusability Assessment

DNNs are novel techniques that have proven to be game-changers in realms such as healthcare, finance, transportation, and technology [2]. Applications include "image and speech recognition" (NLP), recommendation systems, etc. Their ability to learn from data and make predictions is invaluable when dealing with complex problem statements or automating mundane tasks [4], [64]. This section and its connected subsections present the main process of model development and the procedures for reusability assessment shown in Fig. 4.

1) Task selection: Specifying which problem your model will solve, e.g. classification, regression, or clustering.

2) Choosing data types: Depending on the nature of the problem to solve and the availability of sources, one must select the appropriate data types, which could be structured, semi-structured, or unstructured.

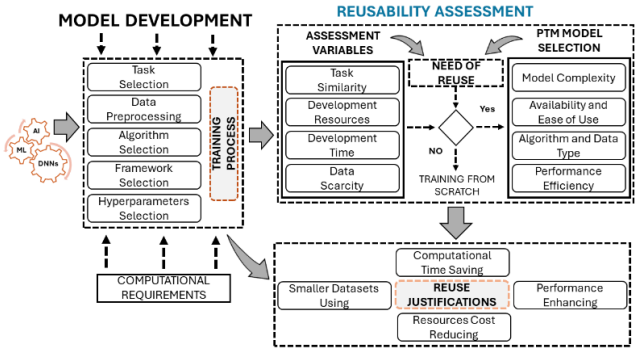


Fig. 4. The development architecture of AI, ML, or DNN-based models, the end-to-end process, and the reusability assessment for the candidate model.

3) Data preprocessing: It includes data cleaning, handling missing values, normalizing (in other words, scaling) features, and encoding them into labels for fitting onto a model.

4) Process steps: The process consists of data collection, preprocessing (sorting), feature selection, model development (training), validation, and testing.

5) Model selection: It involves selecting the appropriate model architecture, such as CNNs for image recognition or RNNs for "sequential data processing".

6) Choose the framework and algorithms: From TensorFlow, PyTorch, or Keras, select the most suitable framework to meet the project's requirements and incorporate algorithms like backpropagation algorithm and gradient descent.

7) Feature extraction: This involves extracting relevant parts of the data using techniques (e.g. principal component analysis), dimensionality reduction, or feature engineering.

8) Model evaluation: Using metrics such as accuracy, precision-recall, and F1 score to determine how efficient an ML algorithm or model is.

9) Reuse and reengineering: We can reuse pre-trained models by fine-tuning them for new, similar tasks, adjusting the model's parameters, or retraining with more data.

Upon completion of candidate construction, the "cross-entropy loss" ($\mathcal{L}_{ce}$) between the "predictions of the target dataset" and the "actual labels" can be calculated by:

$$\mathcal{L}_{ce} = -\sum_{i=1}^{K} t_i \text{Log}(P_i(\mathcal{M},\mathcal{H})) \quad (2)$$

"K" represents "class number" for an intended issue, "$\mathcal{M}$" denotes the "mask," while "$\mathcal{H}$" stands for the "head." Conversely, "$(P_i(\mathcal{M},\mathcal{H}))$" denotes the "prediction a candidate formulates for the class of "i-th" utilizing the "$\mathcal{M}$ mask" and the "head $\mathcal{H}$". Also, the "t_i" signifies the probability of "the class in the single hot representation of the real label," which can be either *0* or *1*. The "classification accuracy" of the "target dataset" is enhanced when the candidate retains a greater number of weights pertinent to the target problem and exhibits a reduced cross-entropy loss. The mask serves to directly calculate the weight retention rate ($\mathcal{L}_{wr}$) as illustrated below:

$$\mathcal{L}_{wr} = \frac{1}{L}\sum_{i=1}^{L} \mathcal{M}[i] \quad (3)$$

L represents the "number of weights in the original mode (3)

"lower weight retention ($\mathcal{L}_{wr}$) rate", suggests that the "candidate model" keeps "fewer weights," whereas "objective function" *O is* defined based on $\mathcal{L}_{ce}$ and $\mathcal{L}_{wr}$:

$$\mathcal{O} = \mathcal{L}_{ce} + \alpha \times \mathcal{L}_{wr} \quad (4)$$

Where α is an empirically determined weighting factor of 1.0.

Importantly, to minimize the $\mathcal{O}$, function, some researchers suggest building a search-based technique that initially focus on find the expected "candidate model that maintains only the weights relevant to the target-related problem" to be solved. The idea is that this "candidate model" has the potential to accomplish the desired "classification accuracy" and "robustness" while maintaining the minimum weights.

G. Model Deployment

In the development section, we highlighted the process of developing a candidate model for a new task. We obtained a candidate model by following all the steps and procedures from the previous stage of development, as shown in Fig. 4. We then did a reusability assessment to make sure it was suitable for the new task and met all its requirements. As a result, the candidate model has been approved for reuse and employment in the new task. This will be achieved through reengineering and reproduction processes through the development of the new target environment, as illustrated in Fig. 5, and the subsequent steps:

1) Infrastructure setup: Deploying the model to the target environment and setting up all back-end infrastructure, such as servers, cloud services, or edge devices.

2) Deploy the model: This is where the trained models are integrated into the target environment, such as a mobile application, web service, or embedded system.

3) Test and validate: Testing the deployed model to ensure that it performs adequately in the target environment.

4) Monitoring and maintenance: Deploy techniques to monitor the model, update it when needed, and maintain ongoing performance support.

The search strategy efficiently explores large models with billions of parameters using a gradient-based method, finding new candidates with smaller objective function values each round and updating the mask accordingly. The updated mask, M1, represents a new candidate with a reduced objective function value, and the process of updating it involves dropping the gradient, as seen below: ξ represents the learning rate:

$$\mathcal{M}' = \mathcal{M} - \xi \times \nabla_{\mathcal{M},\mathcal{H}}\mathcal{O} \quad (5)$$

$$\nabla_{\mathcal{M},\mathcal{H}}\mathcal{O} = \nabla_{\mathcal{M},\mathcal{H}}\mathcal{L}_{ce} + \alpha \times \nabla_{\mathcal{M}}\mathcal{L}_{wr} \quad (6)$$

As a summary, this section outlines the steps involved in deploying the candidate model, which includes setting up the infrastructure for the source model and transforming it into an intermediate model using the intermediate representation (IR) technique, as shown in Fig. 4. The deployment also covers testing, validation, and maintenance after deployment to the target environment. The paper emphasizes the importance of sharing code and models to encourage innovation while also considering the potential consequences of incompatibility. The model's performance efficiency, accuracy, and reliability depend on data selection, preprocessing, and training methods. Successful deployment requires an understanding of all decisions made during the development process.

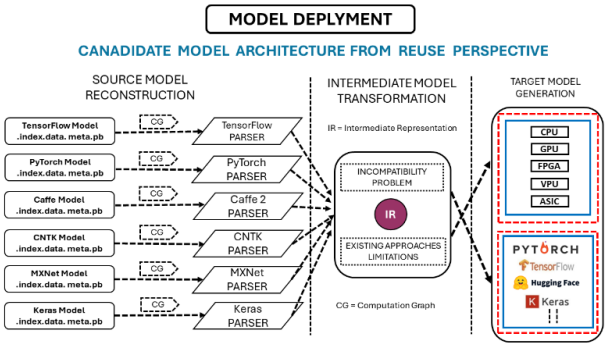


Fig. 5. The deployment architecture of AI, ML, or DNN candidate pre-trained models from source to target environment/s, the end-to-end converting process.

VI. Related Work

In the past decade, specifically the last seven years, there has been a surge in surveys, reviews, and varied studies focused on various applications of machine learning (ML), particularly in complex architectures such as (DNNs) [84], [71], [25]. These systems can potentially enhance multiple business functions and address varied organizational needs [25]. For example, leveraging deep neural networks and related models enables the recommendation of products according to (e.g., previous purchases and audience ratings) [85], [86], image recognition for video surveillance [87], identification of spam and malware emails [88], [89], healthcare applications (e.g. Alzheimer's prediction [78] and cancer prognosis and early detection [90]), among other applications explored by [91], [92]. However, given the novelty of transfer learning, comprehensive survey and review studies have been relatively rare, focusing on providing an overall understanding of this domain [83].

Most notably, relatively little research has been done into the difficulty of compatibility, which can present an insurmountable challenge when attempting to reuse AI systems [55], such as machine learning models and deep neural networks [17]. As far as we know, this systematic review is the first extensive attempt to look at the incompatibility problem through the lens of AI systems reusability. It is an area that has been woefully underexplored, and still, with a deep dive into the nuance of reusability [53] of these cutting-edge technologies, we hope to raise awareness about some of the major issues and factors that come into play [2]. A lack of guidance is available to assist organizations in developing these capabilities [93]. Consequently, our review aims to rectify this gap in literature and serves as a valuable guide to AI system reuse for researchers and practitioners. The integration of ML components into applications faces challenges due to the fragility of algorithms, framework fragmentation, and their susceptibility to changes in data, which can cause prediction shifts over time [94]. Consequently, mismatches between system components hinder the seamless integration of ML capabilities [71]. Thus, [69], [1], [95] Suggested that further research is needed to explore the unique advantages that arise from combining these technologies, particularly considering the increasing availability and complexity of big data characterized by its variety, volume, veracity, volatility, and velocity [96], taking into account understanding the synergy between AI systems and extensive data methods according to [92].

Therefore, this paper aims to fill this gap by comprehensively investigating the state-of-the-art machine and deep learning systems applied in analyzing the bid and complex data [96], including frameworks, models, algorithms, and libraries. It provides an in-depth discussion of current deep neural network technologies, covering their features, categorization, and classification. Additionally, it identifies critical open issues and outlines opportunities for further research in the advancement of big machine deep learning technologies.

VII. The Systematic Review

This review provides a valuable contribution to the research community by offering insights and opportunities for further advancement into the domain of artificial intelligence (AI), particularly advancements in machine learning (ML) technologies and deep neural networks (DNNs). It serves as a guide for researchers and developers aiming to achieve success in evaluating the whole environment of these technologies and networks and related software and hardware components, including the factors that impact the development and deployment process [12], [97]. The first review question (RQ1) seeks to explore the popular AI, ML, and DNN systems and conduct multiple comparisons in various aspects, such as their purposes and goals, tasks, and functions, attributes and characteristics, advantages and disadvantages, similarities and differences, and strengths and weaknesses [98], [99], [100], [101].

The review highlights the role of these systems in preprocessing, training, analyses, storage, and deploying massive and complex data. Multiple comparisons are presented to serve the purpose of this research [98], [39]. The second question (RQ2) specifically aims to identify the open issues and challenges associated with target systems under this study, elaborated in the following sections. The key issues identified include the data and learning complexity, coding from scratch, the dearth of benchmarks, and selecting the proper technology that matches the target task [102], [29], [4], [99], [44], [48], [17]. Through systematic review analysis of the literature, several factors and dimensions affecting these systems were identified, highlighting the significance of the issues closely connected to the primary research problem of incompatibility. [44], [24]. Consequently, understanding these challenges and related factors and dimensions can aid the researchers and the developers in overcoming them, thus achieving the desired results (e.g. prediction, classification, recognition, and detection) [103] (L. Liu et al. 2018b), [99]. The third research question (RQ3) focuses on the main research problem; the findings reveal that interdependencies and compatibility remain open research areas. Hence, further investigation and discussion are needed to find appropriate solutions. The fourth research question (RQ4) addresses the limitations of the existing common empirical approaches that deal with the current interoperability problem and its related challenges [13], [104], [49], [61], [92]. The answer to this question emphasizes the need for a more efficient novel unified method to enhance the compatibility between the source model (e.g., original pre-trained models "PTM") and target model (e.g., the destination model) [102], [48], [105]By addressing these aspects, our authors expect the research community to confidently make strides toward improved compatibility, thus reducing the computational sources' costs and time consumption.

VIII. Methods and Materials

The methods and materials of this study are comprised of three primary stages: the planning of the review, the conducting of the review, the "actual execution of the review," and the reporting of the review. The stages consist of many phases, actions, procedures, measurements, and instruments for achieving the desired findings and outcomes and optimizing the process.

A. The Review Planning

The planning review is a standard step in SLRs that entails two main sets of reviews and procedures. The first set includes formulating research questions, scoping the review, and drafting a review protocol, as shown in Fig. 6. On the other hand, the second set involves identifying the selection of strategy through the development of the (Inclusion/Exclusion) criteria that involve various factors that affect the ML and DNN's performance and reuse process, as presented in Fig. 7. The review process in both sets includes defining objectives, data sources, data extraction, data analysis, and data synthesis.

1) Need of this review: The main goal of this systematic review is to summarize the current evidence on machine learning (ML) techniques and deep neural networks (DNNs), point out gaps in current research on the development and deployment process, and lay the groundwork for new research projects, and explained previously in (Fig. 5 and 6).

2) Research question formulation: The research question formulation stage involves identifying essential elements of domains like AI, ML, and DNNs and addressing methodological aspects such as search, data extraction, and data analysis. The research questions should be specific enough to be feasible within the evaluation scope yet broad enough to make a significant contribution to the field. The research questions include identifying common characteristics and differences, unsolved issues, incompatibility problems, and limitations of existing methods. The goal is to answer these questions while reducing computational processes' high resource costs and time consumption. The review aims to provide a framework for future research and contribute to the fields of machine learning and deep neural networks.

3) Develop the protocol of the systematic review: This study uses a pre-established protocol to avoid researcher bias. It aims to provide a justification for the research, address specific inquiries, and use a systematic approach to identify relevant information. The data extraction strategy identifies variables and methods of interest. The synthesis strategy focuses on addressing primary research inquiries about factors and dimensions impacting the effectiveness of AI, CV, NLP systems, ML techniques, and DNNs. Sub-questions include data synthesis, success definitions, influencing direct factors, indirect factors, and dimension classifications.

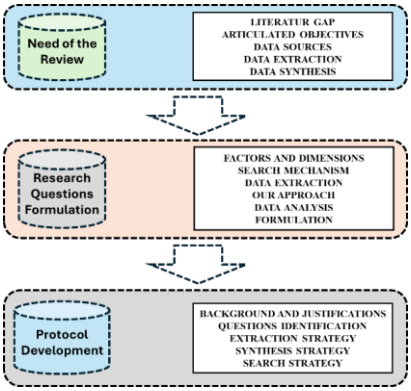


Fig. 6. The first set of review planning includes the need to conduct this review, the research question formulation process, and protocol development.

4) Selection of strategy (inclusion/exclusion criteria): The selection strategy for a review of AI, CV, NLP, ML, and DNN-based systems was based on a series of inclusion and exclusion criteria and derived from research questions and the quality assessment process. The criteria included study focus, empirical studies, evaluation, impact factor, language, theories, publication focus, domain participants, quality control, and replication studies. The emphasis was on English-language studies, focusing on the period from 2015 to 2024 and assessing credible journals.

This review excluded studies that lacked explicit information, publications of low quality, research that focused on artificial intelligence (AI), machine learning (ML), deep-learning/deep neural networks, big and complex data, parallel computing, or technology-based approaches, research that lacked methodology, numerical test findings and analysis, and studies published before 2015 because they were not relevant. Furthermore, we excluded duplicate studies and methods irrelevant to the main research problem.

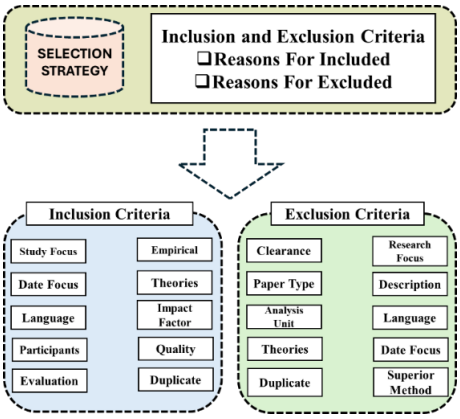


Fig. 7. The second set of review planning involves identifying the strategy selection through the development of inclusion (Inclusion/Exclusion) criteria.

The study focused on addressing the incompatibility observed among AI, CV, NLP, ML, and DNN-based systems, focusing on existing unresolved concerns and constraints. The review included publications from reputable peer-reviewed journals with impact factors and high-quality international scientific conferences.

B. Conducting the Review

The systematic literature review (SLR) is conducted using a well-defined search strategy. Standard electronic databases are the primary source for high-quality primary research, but alternative methods like browsing the internet, seeking advice, and snowballing techniques can also be beneficial. The strategy of search covers selecting the source of data and a search-strings formulating. To determine “search strings”, the authors identify key terms related to machine deep learning software, examine abstracts and titles of selected primary research, and use Boolean operations "OR" and "AND" to construct a comprehensive list of related words. To enrich the search process, we also employed subsequent strings of searches

related to the main keywords and terms. Fig. 8 illustrates a sample of the terms and keywords used.

1) Data sources: The study systematically gathered data from a variety of sources, including highly indexed international scientific conferences and impact-factor journals. Literature selection was based on relevance to AI, ML, and DL (DNNs), as well as related frameworks, models, and algorithms. We used snowballing and backward search techniques to identify additional relevant studies [106]. From a theoretical perspective, the study examined the main concepts, remarkable successes, and achievements of AI, machine learning, and deep neural networks. It also examined related models, algorithms, and frameworks, as well as the related limitations, open issues, and existing challenges. The used theories included Artificial Intelligence, Machine Learning Theory, Deep Learning and Complexity Theory, Computation Theory, Theory of Programming, Transactive Memory Theory, Structured Process Modelling Theory, Coding Theory, Data Science Theory, and Transfer Learning and Reuse Theory.

2) Evaluating eligibility and research quality: It is crucial to evaluate the eligibility and relevance of the primary studies identified in the preceding stage, in addition to the inclusion and exclusion criteria. [107]. As per the instructions outlined in the reference [108], we evaluate the eligibility and quality of each study following the factors, including the significance of the study, the quality of the results and analysis, and potential future research guidelines or discoveries, using the criteria presented in Table I. We evaluate the articles and select studies that exhibit exceptional quality. We have formulated nine questions; each indicated with either (Y)/(Yes) or (P)/(Partly), or (N)/No, to evaluate the eligibility and quality. The questions and their corresponding answers were included. The assigned values for scoring are as follows: Y = 1, P = 0.5, and N = 0. Additionally, each primary study should receive a score ranging from 0 to 13 points. A quality appraisal guarantees that the review will only include the research that is most trustworthy and pertinent to the topic at hand.

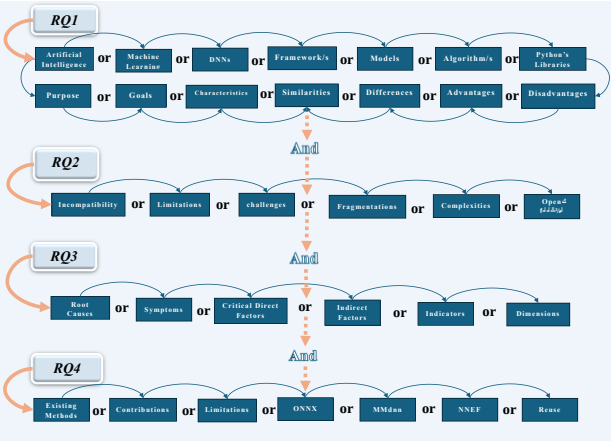


Fig. 8. A list of frequently used terms/strings and keywords on AI, CV, NLP, ML, and DNN-based systems using the Boolean operations "OR" and "AND."

TABLE I. ELIGIBILITY AND QUALITY EVALUATION FORM USING THE "Y-P-N" SCORE

Quality Assessment Questions	Recorded Score
Do the articles provide enough information on "AI, CV, NLP, ML, and DNN" from a theoretical perspective about frameworks, models, and algorithms and their environments?	"Y-P-N"
Does the research study offer a comprehensive comparison between these highlighted techniques?	"Y-P -N"
Does the research paper provide a clear methodology/technique to select proper measurements to evaluate the different aspects of the "AI, CV, NLP, ML, and DNN" domains?	"Y-P -N"
Does the review study present factors, indicators, and dimensions that positively promote the success of "AI, CV, NLP, ML, and DNN" domains?	"Y-P -N"
Do the review papers highlight forward propagation and backpropagation for deep neural networks (DNNs)?	"Y-P -N"
Do the papers efficiently identify the research gaps and well formulate the problem statements?	"Y-P -N"
Do the target studies provide an obvious explanation of the main research problem and causative factors?	"Y-P -N"
Do the target studies provide an explanation and understanding of the related issues challenges and causative factors?	"Y-P -N"
Does the article provide strategies to mitigate or overcome the current DNN issues and challenges?	"Y-P -N"
Does the research article explain and understand the concepts of reuse, transfer learning, and reengineering?	"Y-P -N"
Do the articles provide an explanation, comparison, and understanding of the limitations of the current approaches and techniques?	"Y-P -N"
Are the investigated study outcomes applicable and/or generalizable?	"Y-P -N"
Is the retrieved data appropriately described?	"Y-P -N"
Is the description of inclusion/exclusion/criteria in the study sufficient?	"Y-P -N"

3) The extraction of data: Data extraction in a systematic review is a crucial step in evaluating research potential and combining results. As depicted in Table II, manual data extraction can be laborious and expensive, but sophisticated software can help. In this review, the authors developed a data extraction form to collect primary study information and simplify analysis. Two steps were involved: a preliminary study and a second extraction from a random selection of papers. The data extraction was conducted using tools such as Microsoft Excel, REDCap, and Google Sheets. The intake form included columns for each research question and quality assessment items.

The chosen primary studies provided answers to specific or all the issues raised by the investigations. This systematic and standardized procedure ensures consistency and efficiency in data extraction. Among the results of the 1052 studies, we extracted information that was relevant to each of the four study questions and objectives. Following the extraction of the data, we proceeded to examine and discuss it to accomplish the intended objectives of this SLR exploration.

TABLE II. MAPPING THE INTEREST OF SEARCH TO THE RESEARCH QUESTIONS

Interest of Search	Extracted Data Description
RQ1	"Characteristics, fragmentation, common differences, similarities, advantages, disadvantages, systems, models, libraries, frameworks, methods, tools, techniques, applications, practices, operations, algorithms, and complexities"
RQ2	"Limitations, problems, open issues, current challenges, bottlenecks, processing, complexities, mechanisms, software, fragmentation, diversity, lacking, reusing AI, CV, NLP, ML, DNN, RNN, CNN, (BD), and (DS)."
RQ3	"Preprocessing, learning, training, accuracy, adversarial, robustness, quality, inference time, model size, memory consumption, efficiency, and performance"
RQ4	"Existing methods limitations, transfer learning, reuse, reengineering, conversions, compilers, converters, techniques, accelerators, future trends, and opportunities"
Research Domain	"Systems of AI, CV, NLP, ML, DNN in terms of characteristics, opportunities, and existing challenges.
Utilized Techniques	"Ms. Excel/(spreadsheets), Google/Sheets, and Google/Collab"
Employed Methods	"Whether it is a new method, modified, or hybrid approach"
Relevant Data	"Title/Abstract/Keywords/Authors", "scientific databases/venue"
Publication/ Year	"Included papers from "2015 to 2024" substantially"
Publication/ Type	"Journal, conference, or book chapter"
Study Type	"Whether it is an analysis, survey, or a combination of both"

Deep neural networks (DNNs) have made significant progress in large dataset applications, but they have faced criticism for their ambiguity. This systematic review of AI, machine learning, and DNNs covers conceptualization to dissemination, with a detailed research procedure in six phases. The revised study mapping process follows PRISMA and Kitchenham guidelines, with research questions achieving SLR objectives [109] and kitchenham [110].

C. Phase 1: Broad Preliminary Searching

The preliminary study used Google Scholar to search for relevant material related to AI, CV, NLP, ML, and DNN-based systems. It aimed to derive research questions and establish a variety of search strings and keywords. The preliminary scan of many research papers is conducted for general outlook purposes, facilitating the formulation of the research question. The researcher of this study identified search criteria and selected six eminent scientific sources and submission venues to obtain the most associated studies, identifying 1052 research papers presented in detail in Fig. 9.

D. Phase 2: Identifying Research Criteria to Refine the Search

The search strategy involved selecting keywords/terms related to AI, CV, NLP, ML, and DNN-based systems, Fig. 9. The search string was adjusted for each platform to ensure a systematic search. The review encompassed a variety of criteria and parameters, such as the title, background, review methods, studies included and excluded, results, review questions, discussion, authorship and acknowledgments, citations and references, and appendices.

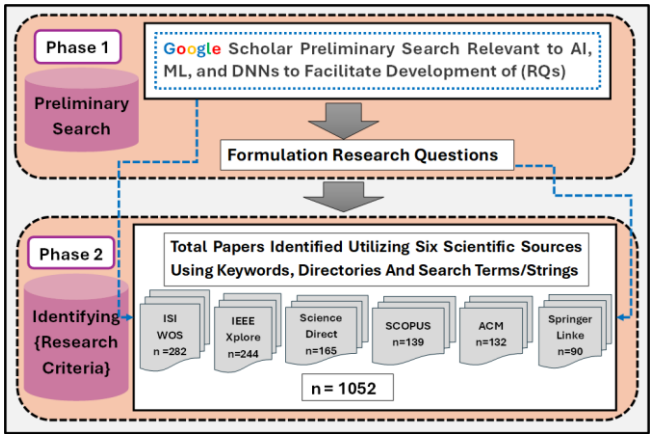


Fig. 9. Phases 1-2. The research question development employed the preliminary search and terms/strings according to the PRISMA and Kitchenham guidelines.

E. Phase 3: Applying the (Inclusion/Exclusion Criteria)

The review phase involved reviewing articles based on inclusion/exclusion criteria, selecting 691 papers related to research questions, while 361 were filtered out, as shown in Fig. 10. The title should be clear and informative, identifying the review's subject and demonstrating its classification as a systematic review.

The review methods should clearly describe the search strategy, the source of data, the study criteria selection, the quality and eligibility assessment methods, and the approach to extracting findings. The discussion section should address the strengths and limitations of evidence, practical implications of findings, research gaps, and areas for future work. The authorship and acknowledgments section should set clear authorship criteria and acknowledge the contributions of individuals or institutions who substantially affected the research but did not meet the author's standards.

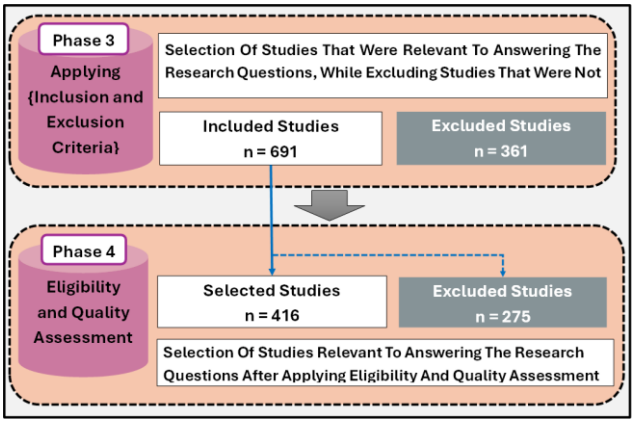


Fig. 10. Phases 3-4. Applying inclusion/exclusion criteria and eligibility/quality assessment according to the PRISMA and Kitchenham guidelines.

F. Phase 4: Evaluating Eligibility and the Quality of Resources

The study critically analyzed 691 selected studies, assessing their suitability and validity. As presented in Fig. 10, we selected 416 studies, ensuring all methodologies were

methodologically appropriate and relevant, and omitted 275 papers due to insufficient eligibility and quality standards.

G. Phase 5: Examining Abstracts and Key Sections

The review process involved a thorough review of 416 papers, focusing on the abstract, primary findings, and conclusion sections, as shown in Fig. 11. The process involved thoroughly examining the articles and identifying those worth further examination. We considered 297 papers for further examination, eliminating 119 due to their unsuitability. The abstract should be 250 words or less long and written in the true style of creative nonfiction.

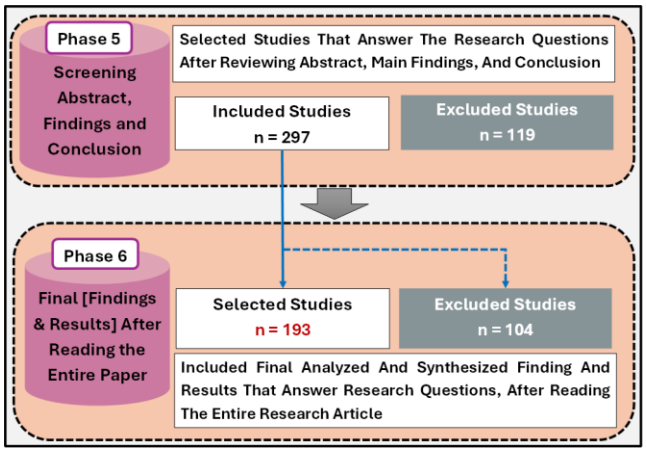


Fig. 11. Phases 5-6. Examining abstracts and key sections and integrating the analyzed selected studies that answer research questions

The findings should include a description of the primary studies, quantitative data, meta-analysis stipulations, a narrative summary of the key findings, and data presentation using tables and diagrams. The conclusion should summarize the main findings, their implications for practice and policy, and suggest suggestions for future research.

H. Phase 6: Further Analysis and Integration

The review comprehensively decoded and comprehended the 297 individual studies in Phase 5, focusing on AI, ML, and DNNs. After presenting the advantages, drawbacks, and existing research issues, the review identified 193 qualitative and quantitative studies. Ultimately, after eliminating 104 studies that were not relevant, the final net tally of the papers is 193. The systematic review report should be formatted as shown in Fig. 11.

IX. FINDINGS

This research is a crucial contribution to advancing deep neural networks, aiming to provide insights and guidelines for advancing the repurposing and adaptation of the DNN model. The systematic review used a six-phase robust approach to search for and extract research on domains of artificial intelligence AI, machine learning ML, deep neural network DNN, and their state-of-the-art techniques.

This study used the search string technique to refine the search area and cross-reference the references and citations of the included studies, thereby identifying additional research, as we can understand from the illustrated information in Table III and the chart in Fig. 12.

TABLE III. A SIX-PHASE ROBUST APPROACH TO CARRYING OUT A SYSTEMATIC REVIEW WAS USED TO SEARCH FOR THE REQUIRED TERMS/KEYWORDS

(Search Source)	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Percentage (%)
(Google Scholar)	Preliminary Search Relevant to AI, ML, and DNNs to Facilitate Development of Research Questions						
ISI-WOS	X	282	193	120	84	54	28%
IEEE	X	244	165	101	72	47	24%
Science Direct	X	165	108	66	50	32	17%
SCOPUS	X	139	87	51	37	25	13%
ACM		132	81	46	32	21	11%
SpringerLink	X	90	57	32	22	14	7%
Total	X	1052	691	416	297	193	100%

A. Phase 1: Groundwork and Theory Building Around Research Questions

Following up on the above essential explanation sections, this phase involved conducting an exploratory Google Scholar search to identify research questions and keywords associated with AI, ML, and DNN and their related applications. The study found that ISI-WoS is the most frequently accessed source of published literature concerning these systems and their associated challenges, followed by IEEEXplore, ScienceDirect, Scopus, ACM, and SpringerLink.

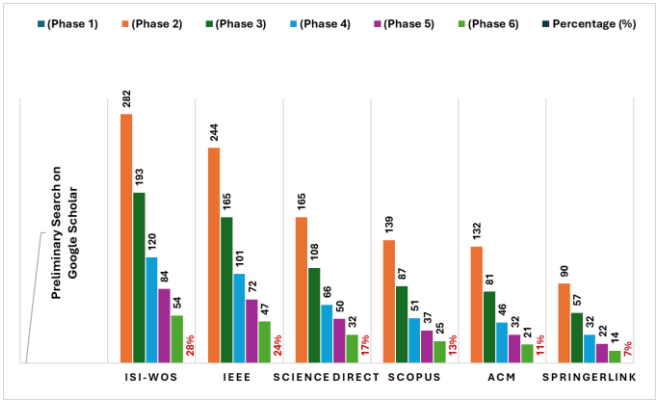


Fig. 12. The intensity of publications for AI, CV, NLP, ML, and DNN-based systems in each scientific database.

B. Phase 2: Identification of Studies with Search Strings: (Narrowing Down the Search Parameters)

Then, we refined the search criteria by defining the acceptable characteristics. This helped us to gather some knowledge of the basics of AI, including ML and DNNs. Deep Neural Networks (DNNs) DNNs are a potent technique that learns multiple layers of data representations or features and yields impressive prediction performances.

The findings identified a total of 1052 publications from six selected scientific sources and databases, simulating a deliberate, non-random selection: ISI-WOS (n=282), IEEE Xplore (n=244), Science Direct (n = 165): SCOPUS (n = 139),

ACM (n=132 papers) Spring-Link ((n=90). These databases allow researchers to search thousands of leading academic journals, periodicals, and conference proceedings.

C. Period of Search

The period focused on in this research included papers from "2015 to 2024" substantially, as depicted in Table IV and Fig. 13.

D. Phase 3: Inclusion/Exclusion Criteria and References Validation

Among the identified studies (n=1052) from the initial search, our reviewers independently assessed inclusions based on predefined criteria. After removing 361 studies that did not meet the requirements (e.g. research questions answers), we were left with 691 relevant to our research questions. We then assessed all eligible studies' references to discover other potentially appropriate reports.

TABLE IV. THE COVERED (TIME-PERIOD) (FROM 2015 TO 2024)

Publication Year	Number of Publications	Percentage (%)
2015	53	5%
2016	69	7%
2017	74	7%
2018	79	8%
2019	88	8%
2020	96	9%
2021	109	10%
2022	121	12%
2023	187	18%
2024	167	16%
Total	1052	100%

E. Phase 4: Evaluation of Eligibility and Adherence Following Reference Review

This phase involved thoroughly evaluating 691 studies and assessing their relevance and quality. We selected 416 studies for their methodological solid foundation and relevance and excluded 275 due to non-compliance with eligibility and quality criteria.

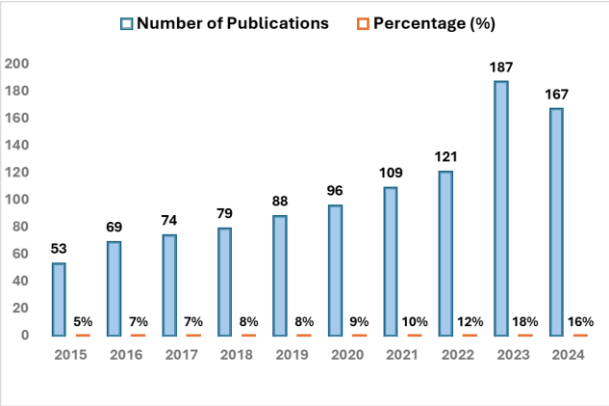


Fig. 13. Overall distribution of AI, CV, NLP, ML, and DNN-based systems from reuse and reengineering perspectives - (Year-Wise).

F. Phase 5: Screening Abstract, Findings, and Conclusion

In this phase, we filtered selection by analyzing abstracts, findings, and conclusions, excluding 119 out of 416 studies for review. This resulted in 297 applicable studies and complete data extraction.

G. Phase 6: Advanced Data Analysis and Answering the RQs

This systematic review concluded with a large set of relevant, high-quality studies that were carefully reviewed, investigated, and synthesized. The study comprehensively analyzed 297 selected studies on AI, ML, and DNNs, excluding 104 irrelevant ones. A structured six-phase process identified and analyzed 193 studies, offering insights to address the study's concerns, as shown in Table V and Fig. 14.

TABLE V. THE ULTIMATE RESULT OF THE REVIEWED PUBLICATION HAS BEEN FOUND AND FILTERED TO BECOME (N = 193)

Publication Year	Number of Publications	Percentage (%)
ISI-WOS	54	28%
IEEE	47	24%
Science Direct	32	17%
SCOPUS	25	13%
ACM	21	11%
SpringerLink	14	7%
Total	193	100%

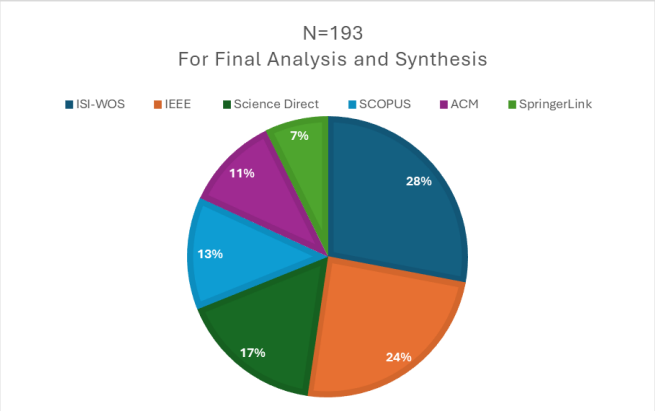


Fig. 14. The scrutinized publication yielded a result of N = 193 after comprehensive analysis and synthesis.

H. Mapping Results and Findings to Research Questions

1) Link each research question to a related dated journal/conference: The systematic review used an iterative design to connect each study to research questions, focusing on RQ1-RQ4. This study developed a new method to classify papers based on the most effectively addressed research questions. Table VI provides details on the mapping process.

TABLE VI. MAPPING PROCESS LISTS THE MOST HIGH-INDEXED REFERENCES PROVIDED ANSWERS TO RESEARCH QUESTIONS

Ref.	Year	Type	Indexing	Cited	Q1	Q2	Q3	Q4
[111]	2019	Journal	ISI/WoS	59	√	√	√	
[99]	2020	Journal	ISI/WoS	58	√	√		√
[6]	2018	Journal	Scie-Dir	56	√		√	
[24]	2020	Journal	Springer	51	√	√	√	
[77]	2021	Journal	Scie-Dir	47	√		√	√
[112]	2020	Journal	Scie-Dir	46	√	√	√	
[113]	2018	Journal	ISI/WoS	40	√	√	√	√
[114]	2019	Journal	ISI/WoS	34		√	√	
[115]	2019	Journal	ISI/WoS	32	√		√	√
[116]	2021	Journal	Scopus	30	√	√	√	√
[117]	2021	Journal	Scie-Dir	30		√	√	
[118]	2018	Conf	IEEE	28	√	√	√	√
[25]	2019	Journal	ISI/WoS	27	√	√		
[119]	2018	Journal	ISI/WoS	26		√		√
[64]	2017	Journal	ISI/WoS	25		√	√	√
[120]	2020	Journal	ISI/WoS	21	√		√	
[73]	2018	Journal	Scie-Dir	20	√	√		√
[121]	2017	Journal	ISI/WoS	19	√	√	√	
[122]	2019	Jr/Cnf	S-D/ACM	18	√		√	√
[123]	2020	Journal	ISI/WoS	17	√	√	√	
[124]	2019	Conf	IEEE	15	√	√		√
[125]	2020	Journal	Scie-Dir	13	√		√	
[126]	2018	Journal	IEEE	12		√	√	√
[127]	2020	Journal	Scie-Dir	11	√	√		
[128]	2019	Conf	IEEE	10	√		√	√
[129]	2020	Journal	ISI/WoS	10	√	√		
[48]	2019	Journal	Springer	9		√	√	√
[130]	2021	Journal	Scie-Dir	9	√	√	√	
[131]	2021	Journal	ISI/WoS	8	√			√
[132]	2018	Conf	IEEE	8	√	√	√	
[133]	2020	Journal	Springer	7	√	√	√	
[134]	2019	Journal	ISI/WoS	7	√		√	√
[135]	2018	Journal	WoS	7	√	√	√	
[136]	2018	Conf	IEEE	6	√	√		√
[137]	2017	Journal	Scie-Dir	6	√	√	√	
[138]	2019	Journal	Scie-Dir	5	√		√	
[139]	2020	Journal	Scie-Dir	5		√	√	√
[44]	2017	Conf	ACM	5	√	√	√	√
[140]	2019	Conf	ACM	5	√		√	
[141]	2018	Conf	WoS/SCIE	5	√	√		√
[142]	2020	Journal/Co	Scopus	4	√	√	√	√
[143]	2017	Conf	IEEE	4	√		√	√
[144]	2020	Journal	Scopus	4	√	√		
[145]	2020	Journal	WoS/SCIE	4	√		√	√
[72]	2019	Journal	WoS/SCIE	4		√	√	√
[146]	2019	Conf	IEEE	4	√	√		
[147]	2020	Journal	Scie-Dir	4		√	√	√
[148]	2018	Journal	IEEE	3	√		√	
[69]	2017	Conf	IEEE	3	√	√		
[149]	2019	Conf	IEEE/ACM	3	√	√	√	√
[13]	2020	Conf	ACM	3	√		√	√
[62]	2021	Conf	IEEE/ACM	3	√	√	√	√
[30]	2023	Conf	IEEE/ACM	2	√	√		√
[49]	2022	Conf	IEEE	2	√		√	
[109]	2022	Journal	WoS	2		√		√
[54]	2023	Symposium	IEEE	2	√		√	
[23]	2023	Conf	IEEE/ACM	2		√	√	√
[150]	2024	Journal	Scopus	2	√		√	
[61]	2024	Journal	WoS	2		√		√
[10]	2024	Journal	Scie-Dir	2	√		√	

2) The most addressed research questions: As depicted in Table VII, the most addressed research question in the reviewed studies was RQ3, which focuses on the characteristics, differences, similarities, advantages, and limitations of AI systems, ML techniques, and DNN networks. However, RQ4, which addresses the current interoperability challenge among AI environments, was the least addressed, with only 35 papers providing answers. The research papers that answered RQ4 were average and insufficient, highlighting the need for further research into AI system issues. Despite the fewer papers addressing RQ4, some notable contributions include new methods for distributed training and parallel computing.

TABLE VII. PERCENTAGE (%) OF STUDIES COVERING EACH RESEARCH QUESTION

Question No	Question-Text	Number of Publications	Percentage (%)
"RQ1"	"What are the main characteristics and "common differences, similarities, advantages, and limitations" of AI systems, ML techniques, and DNN networks regarding purpose, tasks, and operations?"	54	28%
"RQ2"	"What are the key unsolved problems and bottlenecks that AI computational environments face?"	45	23%
"RQ3"	"What are the relevant "factors and dimensions" that impact the accuracy, adversarial robustness, memory consumption, and inference time of AI, ML, and DNN tasks?"	59	31%
"RQ4"	"What convenient approaches can be developed to address the existing issues and reduce the computation process's high resource cost and time consumption?"	35	18%
Total	Four questions	193	100%

I. Characteristics of (AI) Systems, ML Technologies, and DNNs and Answering RQ1

This systematic literature review comprehensively compares popular machine learning and deep learning frameworks (e.g. TensorFlow, Caffe-Caffe2, Torch/PyTorch, Theano, MXNet, and CNTK) and their connected models and algorithms, as illustrated in Table VIII, towards answering RQ-1. It highlights their relevant characteristics in various aspects and demonstrates the significant disparity in compatibility

between frameworks, with excellent compatibility for models from the same ecosystem. However, compatibility diminishes when combining models from different ecosystems, emphasizing the importance of considering framework compatibility during model development.

TABLE VIII. A COMPREHENSIVE COMPARISON OF (AI) SYSTEMS, ML TECHNOLOGIES, AND DNNS AND ANSWERING [RQ1]

Framework Factors	Caffe/Caffe2	TensorFlow	Theano	Torch/PyTorch	CNTK	MXNet
Founder	Learning-Center & Berkeley 2013/caffe 2017/caffe2	Google2015	Montréal-Uni (2007)	(Facebook)2016	Microsoft Research 2016	Apache Foundation 2015
Purpose	Image processing & on both CPU and GPU	High-performance for ML, DL, CPU, GPU, and TPU	Python offers GPU for arrays.	A Python-based for DNNs	For DL, graph representation	A DL for DNNs.
Interface	Python, MATLAB	Python (Keras), C/C++, Java, Go, &R	Python	Python, C++, & Julia	Python, C++,	C++, Python, Julia, Matlab,
Core language	C++	C++	Python	Python, C, C++, CUDA	C++	Small C++ core library
Platform	Linux, Mac OS X, Windows	Linux, Mac OS X, Windows	Cross-platform	Linux, macOS, Windows	Windows, Linux	Linux, Mac OS X,
CUDA Support	Yes	Yes	Yes	Yes	Yes	Yes
Multi-Threaded -CPU	Yes (BLAS)	Yes (Eigen)	Yes	Yes (widely Used)	Yes	Yes (OpenMP)
Pretrained Models	Yes	Yes	Partially	Yes	Yes	Yes
Multi-GPU	Yes (only data parallel)	Yes (Most flexible)	Not perfect	Yes	Yes	Yes
Compatibility (within the same environment)	Yes	Yes	Yes	Yes	Yes	Yes
Compatibility in	No	No	No	No	No	No
Advantages	"Fast, Efficient for CNNs"	"Large Community, Flexible"	"Symbolic, Good for Research"	"Imperative, Pythonic"	"Production-ready"	"Scalable, Efficient"
Disadvantages	"Limited framework support"	"Can be complex."	"Slow execution"	"Limited community"	Win limited	"Learning curve"

J. Challenges and Open Issues and Answering of RQ2

1) Overview: The systematic review highlights several challenges facing AI, ML, and deep neural networks (DNNs) and their related frameworks, models, and algorithms. These issues include system fragmentation, complexity, interpretability, bias, security, privacy, and a lack of benchmarks. System fragmentation affects code and model reusability, while complexity involves data, model size, algorithm complexity, high dimensionality, and characteristics of tensor data used in DNNs. Understanding how machine-learning techniques and DNNs make decisions like black boxes makes interpretability challenging. Bias can lead to unfair and discriminatory outcomes, while security is vulnerable in specific fields like autonomous vehicles and financial institutions.

2) Common Challenges

- ***Interpretability:*** Understanding how machine learning (ML) and deep neural networks (DNNs) make decisions as black boxes can be challenging. Trusting the procedures throughout the learning and training process can be challenging without reliable indicators until each iteration, epoch, and batch size yields the desired production results (e.g. prediction accuracy before and after model deployment) [151].
- ***Bias:*** Machine learning and deep learning models can be biased, which can lead to unfair and discriminatory outcomes [152].
- ***Security:*** Machine learning systems, especially in some fields such as autonomous vehicles and financial institutions, can be vulnerable to security and safety attacks [153], [154], [155].
- ***Privacy:*** The intelligent systems discussed in this review can collect and store large amounts of personal data, which raises privacy concerns [154], [156]. This issue is more challenging when dealing with deep neural (DNN) models, as these models need a huge amount of data that is not readily provided, especially in applications that require lots of sensitive data [156]. Moreover, Mendelson and Avi have demonstrated that privacy has been a hot topic, particularly regarding the need for privacy protection [154]. Since then, that has translated into new legislation affecting our rights regarding how they get data collected on them. Furthermore, the authors have implemented new mechanisms to guard against the potential misuse of this information. However, the current initiatives are neither adequate nor sufficient.
- ***Lack of benchmarks:*** Benchmarks are essential for comparing the performance of different AI environments and identifying areas where improvement is needed. According to [29], there is a lack of standard benchmarks for evaluating the machine-learning perf (ML) and DNNs performance for the diver's data, especially "big data analytics" [157], [158]. The need for more benchmarks within these technologies is a particularly significant challenge due to the need for standard benchmarks [159]. Furthermore, several surveys indicate that few discussions have been held about the

community's efforts to implement deep learning benchmarking and the gaps between what has been achieved and what still needs to be added [160]. It would be a valuable contribution to the field of research, and developers would pay more attention to this challenge for industry-standard benchmark development [159], [29].

- ***Difficulty in choosing*** *between the machine learning techniques:* In general, the selection of the proper machine learning and deep neural methods for a given task can be a challenge [44], [99]. Based on the analyzed studies, the systematic review identified factors that affect the performance of machine learning and deep neural systems, such as the type of selected data, the quality and quantity of the training data, the kind of algorithm, model selection, the computational resources availability, and the appropriate hyperparameters and fine-tuning for the intended task [48]. Moreover, the findings have expounded various implications for researchers and practitioners. Researchers should develop new methods to address the open issues and challenges identified in the analyzed and synthesized SLR. On the other hand, practitioners should be aware of the involved factors that affect the computation performance and factors that should be applied carefully to select the proper techniques (e.g. Framework, pre-trained model, algorithm, and other software components) and hardware requirements (e.g. Computer device, memory size, and accelerators such as CPUs GPUs, and TPUs) for their given specific needs.

3) Learning and Training Challenges in DNNs

- ***Insufficient Training Data:*** Low training data or batch size can hinder stable, generalizable solutions. Ensure that there is sufficient data for the local minimum cost function.
- ***Non-deterministic data:*** Breaking the deterministic relationship between training inputs and target outputs can hinder training convergence.
- ***Wrong Learning Rate:*** Small or large learning rates can slow or make training unstable.
- ***Vanishing Gradients:*** Gradients in particular layers can become small, slowing or stopping training. For other solutions, use different activation functions.
- ***Exploding Gradients:*** Gradients can become large and cause instability, limiting their weight and bias change rate.

K. The Critical Direct and Indirect Factors Impact DNNs and Answering RQ3

This study examined various factors that affect the DNN-based system's performance and robustness. These factors are critical, direct, and indirect, as explained in Fig. 15.

Direct Factors	Definition	Indirect Factors	Description
Fragmentation	Each system has its own features and characteristics.	NCHW/NHWC Formats	Different layouts of data in the memory hierarchy. The algorithm must be transposed to lead to poor performance.
Structure	Each system has its own structural design.	Data Noise	Mislabeled cases, redundant data, borderlines, or errors in the values of attributes.
Complexity	Model Complexity, Algorithm Complexity, Data/Tensor, and Tensor Layout (Curse of Dimensionality).	Outliers	It is a broader concept that contains discordant data
Purpose	Framework Goal, Model Goal, Algorithm Goal, Task Type (e.g., Prediction, Classification, Detection, or Recognition).	Data Size	Data size: (Not sizable).
Programing Language	API, core language, and Frontends/Interface (C++, Python, Julia, Matlab, JavaScript, Go, R, Scala, or Perl)	Imbalance Data	One or some of the classes greater than others result in biased outcomes.
Accelerators	CPUs, GPUs. TPUs, FPGA, VPU, and ASIC	Missing Features	Variables lack information points, resulting in harming accuracy and model dependability.

Fig. 15. Critical direct and indirect factors impact the DNNs reuse and relevant reengineering challenges (n1=6 direct, n2=6 indirect factors).

The direct factors include data quality, structural design, complexity, creation purpose, programming language, and accelerators. On the other hand, indirect factors such as noise, outliers, and imbalances can impede DNNs' ability to learn program scale and dependencies. By carefully designing and addressing these indirect factors, DNNs can receive training on high-quality data for more accurate and reliable results. We can use deep learning accelerators like GPUs, TPUs, FPGAs, VPUs, and ASIC chips to meet real-time requirements. Therefore, our results demonstrate that researchers and developers should focus on developing more efficient methods to address the highlighted challenges. At the same time, practitioners should be aware of factors affecting computation performance and carefully select the proper techniques and hardware requirements for their specific needs.

L. Existing Limitations on DNNs and Answering RQ4

1) Overview: A potential way for companies and the computer industry to amortize costs is by reusing and re-purposing DNNs. However, reusing DNN-based systems face challenges that affect their reuse and reengineering performance, efficiency, and robustness. The direct and indirect factors behind these challenges have been highlighted in previous sections. The structural problem of non-interoperability among DNN-based systems is one of these obstacles, along with a lack of technical skills and engineering methods [3].

In deep neural cutting-edge systems, interoperability refers to the ability of one software component (with which models are associated) to exchange algorithms and data with another [80]. Fig. 16 illustrates the various paths in developing and deploying a model on hardware or a target environment.

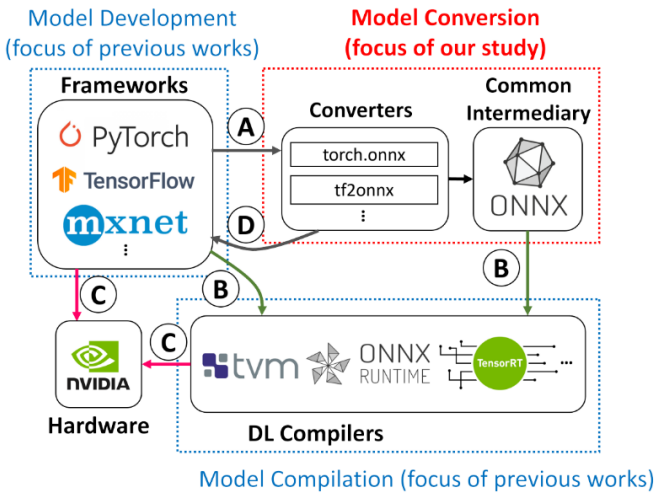


Fig. 16. The paths among model development, hardware deployment, DL Compilers, and model conversion [161].

Interoperability lets models be reused across frameworks and deployment scenarios. While (A) signifies model conversion into a standard intermediary format, letter (B) indicates the compilation process. On the other hand, (C) represents model deployment, and (D) describes transforming a model into a framework according [161].

2) Failure of conversion: This software ecosystem mainly includes many dedicated frameworks for deep learning [162], [163] and compilers [164], [105]. The proliferation of these different tools for development and deployment represents a significant obstacle since moving algorithms across frameworks can be difficult. Furthermore, a compiler that understands models from various frameworks is very heterogeneous. Consequently, frameworks such as open neural network exchange (ONNX) are developed to play an essential role in mediating between deep learning software [165].

However, translation errors during conversion can result in distortions and affect the quality of the model, potentially leading to a degradation in prediction, impacting deployment efficiency, and ultimately causing conversion failure.

3) Failure symptoms and root causes: Even with the development of modern interoperability solutions such as "ONNX, MMdnn, CoreML, and NNEF," there remains a knowledge gap regarding the failure type of deep learning model conversion errors. This lack of knowledge exposes a significant risk with various interoperability technologies built into the design.

The study findings have divided the failure symptoms into two categories: The main common failure symptoms and relevant practical root causes of reusing and reengineering DNN-based techniques, as elaborated in Fig. 17. The second category is failure symptoms that arise from the structural and procedural root causes when reusing and reengineering DNN systems, as presented in Fig. 18.

- Main and Common Failure Symptoms

Model Size: In memory-constrained environments, pre-trained models (PTM) with large sizes are challenging to deploy and use.

Inference Time: Reusing the entire set of pre-trained models can increase the inference time, affecting real-time applications.

Defect Inheritance: Reusing models without a code review may result in inheriting flaws from the original model, which could pose security risks in the new system. On the other hand, AI models are famously brittle to modest input data perturbations, allowing attackers to create adversarial examples for malevolent assaults.

Poor Performance: Compared to regression testing or hardware performance, the model converter's time and memory costs are substantially higher than developer/user expectations.

Failure Symptom	Definition	Failure Root Cause	Definition
Model Size	In memory-constrained environments, pre-trained models (PTM) with large sizes are a significant challenge to deploy and use.	Reusing the entire Model	The entire model is being reused, even though only a small portion of its functionalities (mainly labels) are necessary.
Inference Time	Reusing only a small portion of the functions provided by the model can lead to overhearing and unwanted computational costs.	Training the entire set of pre-trained models	Reusing the entire set of pre-trained models can increase the inference time, affecting real-time applications.
Defect Inheritance	Inheriting flaws from the original model could pose security and robustness risks in the new system (Weakness Inheriting).	AI models are mostly brittle to modest input data perturbations,	Reusing models without a code review.
Poorly Performance	Compared to regression testing or hardware performance, the model converter's time and memory costs are substantially higher than developer/user expectations.	Algorithmic Error	Optimization: The model converter is experiencing an issue with optimizations, including the elimination of inactive code or incorrect fusing. Tracing: Tracking DNN poses a challenge. This phenomenon is observed in frameworks that employ dynamic models. Dynamic models must be tracked to identify the nodes in the computational graph prior to the conversion.

Fig. 17. The main common failure symptoms and relevant practical root causes of reusing and reengineering DNN techniques (n1=4 symptoms, n2=4 causes).

- Structural failure symptoms

In General, interoperability relies on deep neural network (DNN) conversion methods such as "NNEF, ONNX, MMdnn, and CoreML." While primarily used for deployment and framework conversion, these methods suffer from crashes and performance degradation. This study demonstrated three structural failure types, as shown in Fig. 18:

Crashes and wrong model behavior: because of type of problem and incompatibility issues.

Wrong model selection: because of tensor shape mismatches during transformation, inference, or layout conversion.

Build failures: These occur because of testing issues like missing or broken tests. These issues primarily impact converters' graph translation and optimization stages. This study's findings align with existing research, and further analysis revealed model characteristics linked to converter failures.

Failure Symptom	Definition	Failure Root Cause	Description
Crash	Typically, a report of error is generated when the model converter terminates unpredictably during the conversion process.	Incompatibility	Internal: The API evolution has resulted in API compatibility issues within a model converter. External: There are API compatibility issues between a model converter and third-party libraries (e.g., ONNX, MMdnn, NNEF, TensorFlow, ONNX Run-time).
		Problem Type	Node: A model converter employs a computational network, with nodes representing DL operators such as convolution and pooling and edges representing tensors. Every node receives a variable number of tensors and outputs one. These challenges involve node transformation. Tensor: A tensor is a matrix with several dimensions that consists of elements of the same data type. This applies to the various classifications of tensors. Conventional: Conventional software systems also use customary factors. This subcategory addresses traditional variables in various forms.
Wrong Model	model converter exhibits unanticipated behavior without crashing, resulting in an incorrect—intermediate or final result.	Tensor Shape	This affects the shape of a node's input and output tensors during shape matching, transformation, inference, layout, etc.
Build Failure	Installation of model converter or dependencies fails.	Testing	Problems associated with software tests, including unit tests. This encompasses "flaky" tests, missing tests, broken tests, or novel tests.

Fig. 18. Structural failure symptoms arise from the structural and procedural root causes of DNN reuse (n1 =3 structural symptoms, n2 =4 structural causes).

M. Opportunities

Merging innovation with interoperability of deep neural models:

This study suggests that DNN applications need additional interoperability research to understand how pre-trained models make decisions to improve transparency and accountability. A deeper understanding of model converter failure modes may also lead to improved solutions. Scaling these new technologies may ease existing approaches' limits, enabling scale-up for any emerging model and reducing resource and time costs. Our review study (RQs) and relevant critical direct/indirect factors, failure symptoms, and root causes can help researchers and developers overcome limitations in ML and deep neural network model development and deployment from reuse inspiration. Currently, the study team is developing an innovative and promising approach that could overcome the incompatibility and related reuse issues to enhance the re-purposing capabilities of the DNN's domain and practice.

X. Conclusion and Future Trends

Deep neural networks (DNNs) have revolutionized the fields of machine learning (ML) and artificial intelligence (AI), enabling the creation of various models and algorithms capable of autonomously solving complex tasks. However, the development of these techniques faces significant challenges. This paper comprehensively reviews deep neural models that address the characteristics, challenges, open issues, and critical factors influencing ML and DNN technologies. It guides researchers and developers in achieving successful DNN models from reuse inspiration by evaluating factors impacting the reuse and related learning pipeline process. It carefully examines the symptoms of failure in existing reuse methods and the causes that usually lead to this failure. The SLR examined four research questions to comprehend their characteristics, challenges, influencing factors, and strategies for resolving limitations, particularly incompatibility and associated issues. By analyzing the first 193 research papers, the study identified significant challenges related to software incompatibility within DNNs. Other open issues include performance degradation, complexity, vulnerability to adversarial attacks, difficulty in code reengineering and conversion, a lack of standardized benchmarks, and system selection. The review reveals how these issues hinder parallel computing efficiency by increasing development, deployment, and computing time, complicating task organization, and raising costs. Additionally, goal mismatches can decrease computing accuracy, making training and performance complex. This study contributes to understanding DNN as a black box and benefits various fields, including AI, ML methods, and DNN-based systems in performance (e.g., accuracy, inference time, memory consumption) and robustness towards multiple adversarial attacks. The study suggests that DNN applications require further research on interoperability, model converter failure modes, and scaling technologies to improve solutions and reduce costs.

Future work focuses on reducing model size computational cost without sacrificing performance efficiency, whether for basic DNN models (e.g., pre-trained models) or LLMs using unique techniques (e.g. "Low-Rank-Adaptation "Lora" and Quantized Lora "QLora," model compression, and pruning) for optimal converted model size and inference time). Furthermore, during future work procedures, the team ensures that various adversarial attacks, especially in finance, healthcare, and autonomous applications, are avoided, including data security, privacy, veracity, and benchmarking challenges.

Acknowledgment

The authors gratefully acknowledge all contributors whose valuable insights and suggestions significantly improved this work. We confirm that no funding was received for this work and declare no conflicts of interest.

References

[1] Y. Duan, J. S. Edwards, and Y. K. Dwivedi, "Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda," Int J Inf Manage, vol. 48, no. February, pp. 63–71, 2019, doi: 10.1016/j.ijinfomgt.2019.01.021.

[2] J. Liu et al., "Usability Study of Distributed Deep Learning Frameworks For Convolutional Neural Networks," 2019. [Online]. Available: https://caffe2.ai/

[3] J. C. Davis, P. Jajal, W. Jiang, T. R. Schorlemmer, N. Synovic, and G. K. Thiruvathukal, "Reusing Deep Learning Models: Challenges and Directions in Software Engineering," Apr. 2024, [Online]. Available: http://arxiv.org/abs/2404.16688

[4] Z. Wang, K. Liu, J. Li, Y. Zhu, and Y. Zhang, "Various Frameworks and Libraries of Machine Learning and Deep Learning : A Survey," Archives of Computational Methods in Engineering, no. 0123456789, 2019, doi: 10.1007/s11831-018-09312-w.

[5] J. Liu et al., "Usability Study of Distributed Deep Learning Frameworks For Convolutional Neural Networks," ACM ISBN 978-x-xxxx-xxxx-x/YY/MM. In Proceedings of the Deep Learning Day at SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'18)., no. 1, p. 9, 2018, [Online]. Available: https://caffe2.ai/

[6] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, "A survey on deep learning techniques for image and video semantic segmentation," Applied Soft Computing Journal; Publisher: ELSEVIER , RADARWEG 29, AMSTERDAM, NETHERLANDS, 1043 NX, vol. 70, pp. 41–65, 2018, doi: 10.1016/j.asoc.2018.05.018.

[7] M. Hilbert, "Big Data for Development: A Review of Promises and Challenges," Development Policy Review peer-reviewed academic journal Impact factor:1.093 2019 SJR, vol. 34, no. 1, pp. 135–174, 2016, doi: 10.1111/dpr.12142.

[8] S. Dobilas, “Transposed Convolutional Neural Networks — How to Increase the Resolution of Your Image,” pp. 1–24, 2022.

[9] I. Tecnol, “Performance Comparison of Deep Learning Frameworks in Image Classification Problems using Convolutional and Recurrent Networks,” In2017 IEEE Colombian Conference on Communications and Computing (COLCOM) 2017 Aug 16 (pp. 1-6). IEEE. 978-1-5386-1060-2/17/$31.00 ©2017 IEEE, no. In2017 IEEE Colombian Conference on Communications and Computing (COLCOM) 2017 Aug 16 (pp. 1-6). IEEE., p. 6, 2017.

[10] J. Just, “Natural language processing for innovation search – Reviewing an emerging non-human innovation intermediary,” Technovation, vol. 129, Jan. 2024, doi: 10.1016/j.technovation.2023.102883.

[11] A. Khang, V. Abdullayev, O. Hrybiuk, and A. K. Shukla, “COMPUTER VISION AND Al-INTEGRATED IOT TECHNOLOGIES IN THE MEDICAL ECOSYSTEM EDITED BY,” 2024.

[12] J. Rezaeenour, M. Ahmadi, H. Jelodar, and R. Shahrooei, “Systematic review of content analysis algorithms based on deep neural networks,” Multimed Tools Appl, pp. 17879–17903, 2022, doi: 10.1007/s11042-022-14043-z.

[13] Y. Liu et al., “Enhancing the interoperability between deep learning frameworks by model conversion,” Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, p. 10, 2020, doi: 10.1145/3368089.3417051.

[14] B. Jan et al., “Deep learning in big data Analytics: A comparative study,” Journal of Computers and Electrical Engineering Science Direct, vol. 75, no. September 2018, pp. 275–287, 2019, doi: 10.1016/j.compeleceng.2017.12.009.

[15] A. Shatnawi, G. Al-Bdour, R. Al-Qurran, and M. Al-Ayyoub, “A comparative study of open source deep learning frameworks,” 2018 9th International Conference on Information and Communication Systems, ICICS 2018, vol. 2018-Janua, no. April, pp. 72–77, 2018, doi: 10.1109/IACS.2018.8355444.

[16] M. P. Reyes et al., “A Survey on Deep Learning,” ACM Comput Surv, vol. 51, no. 5, pp. 1–36, 2018, doi: 10.1145/3234150.

[17] J. Nilsson, System of Systems Interoperability Machine Learning Model System of Systems Interoperability Machine Learning Model. Dept. of Computer Science and Electrical Engineering Lule˚a University of Technology Lule˚a, Sweden: Printed by Luleå University of Technology, Graphic Production 2019 ISSN 1402-1757 ISBN 978-91-7790-458-8 (print) ISBN 978-91-7790-459-5 (pdf) Luleå 2019 www.ltu.se, 2019. [Online]. Available: www.ltu.se

[18] Jason . Brownlee, “A Gentle Introduction to Tensors for Machine Learning with NumPy,” pp. 1–17, 2018.

[19] Facebook Open Source, “Caffe2 Open Source Brings Cross Platform Machine Learning Tools to Developers,” Https://Caffe2.Ai/, pp. 3–6, 2017, [Online]. Available: https://caffe2.ai/blog/2017/04/18/caffe2-open-source-announcement.html

[20] J. Li et al., “Feature selection: A data perspective,” ACM Computing Surveys Journal ISI:Science Citation Index Expanded, vol. 50, no. 6, 2017, doi: 10.1145/3136625.

[21] Y. E. Wang, C.-J. Wu, X. Wang, K. Hazelwood, and D. Brooks, “Exploiting Parallelism Opportunities with Deep Learning Frameworks,” ACM Digtal librar ACM Trans. Archit. Code Optim., Vol. 18, No. 1, Article 9, Publication date: December 2020., vol. abs/1908.0, 2020.

[22] Y. You, J. Demmel, K. Keutzer, Z. Zhang, and C. Hsieh, “Fast Deep Neural Network Training on Distributed Systems and Cloud TPUs,” IEEE Transactions on Parallel and Distributed Systems, vol. 30, no. 11, 2019, doi: 10.1109/TPDS.2019.2913833.

[23] B. Qi, H. Sun, X. Gao, H. Zhang, Z. Li, and X. Liu, “Reusing Deep Neural Network Models through Model Re-engineering,” Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.00245

[24] Z. Wang, K. Liu, J. Li, Y. Zhu, and Y. Zhang, “Various Frameworks and Libraries of Machine Learning and Deep Learning: A Survey,” Archives of Computational Methods in Engineering, no. 0123456789, p. 24, 2019, doi: 10.1007/s11831-018-09312-w.

[25] Y. Wu et al., “A Comparative Measurement Study of Deep Learning as a Service Framework,” IEEE Trans Serv Comput, no. Dl, p. 15, 2019, doi: 10.1109/TSC.2019.2928551.

[26] L. Ruthotto, S. J. Osher, W. Li, L. Nurbekyan, and S. Wu, “A machine learning framework for solving high-dimensional mean field game and mean field control problems,” Proceedings of the National Academy of Sciences, PNAS April 28, 2020 117 (17) 9183-9193; first published April 9, 2020, vol. 117, no. 17, p. 11, 2020, doi: 10.1073/pnas.1922204117.

[27] W. Jiang, G. Zeng, S. Wang, X. Wu, and C. Xu, “Application of Deep Learning in Lung Cancer Imaging Diagnosis,” J Healthc Eng, vol. 2022, 2022, doi: 10.1155/2022/6107940.

[28] P. Skryjomski, B. Krawczyk, and A. Cano, “Speeding up k-Nearest Neighbors classifier for large-scale multi-label learning on GPUs,” Neurocomputing Journal Publisher: ELSEVIER , RADARWEG 29, AMSTERDAM, NETHERLANDS, 1043 NX; ISI: Science Citation Index Expanded, vol. 354, 2019, doi: 10.1016/j.neucom.2018.06.095.

[29] F. Bajaber, S. Sakr, O. Batarfi, A. Altalhi, and A. Barnawi, “Benchmarking big data systems: A survey,” Comput Commun, vol. 149, pp. 241–251, 2020, doi: 10.1016/j.comcom.2019.10.002.

[30] W. Jiang et al., “An Empirical Study of Pre-Trained Model Reuse in the Hugging Face Deep Learning Model Registry,” Mar. 2023, [Online]. Available: http://arxiv.org/abs/2303.02552

[31] R. A. Hazarika, A. Abraham, D. Kandar, and A. K. Maji, “An Improved LeNet-Deep Neural Network Model for Alzheimer’s Disease Classification Using Brain Magnetic Resonance Images,” IEEE Access, vol. 9, pp. 161194–161207, 2021, doi: 10.1109/ACCESS.2021.3131741.

[32] J. Zhang, X. Yu, X. Lei, and C. Wu, “A Novel Deep LeNet-5 Convolutional Neural Network Model for Image Recognition,” Computer Science and Information Systems, vol. 19, no. 3, pp. 1463–1480, Sep. 2022, doi: 10.2298/CSIS220120036Z.

[33] R. U. Khan, X. Zhang, R. Kumar, and E. O. Aboagye, “Evaluating the performance of ResNet model based on image recognition,” in ACM International Conference Proceeding Series, Association for Computing Machinery, Mar. 2018, pp. 86–90. doi: 10.1145/3194452.3194461.

[34] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions,” Oct. 2017, [Online]. Available: http://arxiv.org/abs/1610.02357

[35] C. Wang et al., “Pulmonary image classification based on inception-v3 transfer learning model,” IEEE Access, vol. 7, pp. 146533–146541, 2019, doi: 10.1109/ACCESS.2019.2946000.

[36] D. F. and A. V. Hussam Qassim, “VGG16 - Residual Squeeze,” 2018.

[37] [37] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Sep. 2017, [Online]. Available: http://arxiv.org/abs/1409.1556

[38] S. Sztwiertnia, M. Grübel, A. Chouchane, D. Sokolowski, K. Narasimhan, and M. Mezini, Impact of programming languages on machine learning bugs, vol. 1, no. 1. Association for Computing Machinery, 2021. doi: 10.1145/3464968.3468408.

[39] C. Luo, X. He, J. Zhan, L. Wang, W. Gao, and J. Dai, “Comparison and Benchmarking of AI Models and Frameworks on Mobile Devices,” arXiv preprint arXiv:2005.05085. Cornell University Journal, vol. abs/2005.0, 2020, [Online]. Available: https://www.semanticscholar.org/paper/3858ccebd64b63e27a8cbc2d4d4498e063c05688

[40] A. Chapman et al., “Dataset search: a survey,” in VLDB Journal, Springer, Jan. 2020, pp. 251–272. doi: 10.1007/s00778-019-00564-x.

[41] N. Sharma, V. Jain, and A. Mishra, “An Analysis of Convolutional Neural Networks for Image Classification,” in Procedia Computer Science, Elsevier B.V., 2018, pp. 377–384. doi: 10.1016/j.procs.2018.05.198.

[42] Y. Wang et al., “Benchmarking the Performance and Energy Efficiency of AI Accelerators for AI Training,” Proceedings - 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing, CCGRID 2020, pp. 744–751, 2020, doi: 10.1109/CCGrid49817.2020.00-15.

[43] Y. Wang, G. Y. Wei, and D. Brooks, “Benchmarking TPU, GPU, and CPU platforms for deep learning,” arXiv preprint arXiv:1907.10701. Coenell University, vol. 4, 2019.

[44] L. Nguyen, P. Yu, and M. Chowdhury, “No!: Not Another Deep Learning Framework,” Proceedings of the Workshop on Hot Topics in Operating Systems - HOTOS ’17, Whistler, BC, Canada, May 08-10, 2017, https://doi.org/10.1145/3102980.3102995. ACM, vol. Part F1293, no.

May 08-10, 2017, pp. 88-93 (6 Pages), 2017, doi: 10.1145/3102980.3102995.

[45] T. Zhao and X. Huang, "Design and implementation of DeepDSL: A DSL for deep learning," Comput Lang Syst Struct, vol. 54, pp. 39–70, 2018, doi: 10.1016/j.cl.2018.04.004.

[46] W. Xiang, "APPLICATIONS OF DEEP LEARNING IN LARGE-SCALE OBJECT DETECTION AND SEMANTIC SEGMENTATION," ProQuest Number:13855813, no. March, p. 142, 2018.

[47] V. Kovalev, A. Kalinovsky, and S. Kovalev, "Deep Learning with Theano, Torch, Caffe, TensorFlow, and Deeplearning4J: Which One Is the Best in Speed and Accuracy? Recognition of multi-drug resistant tubercolusis based on CT and X-ray image analysis View project UAV: back to base problem View project," XIII International Conference on Pattern Recognition and Information Processing; SPRINGER, pp. 99–103, 2019, [Online]. Available: http://imlab.grid.by/

[48] G. Nguyen et al., "Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey," Artificial Intelligence Review (2019) 52:77–124 Springer, vol. 52, no. 1, pp. 77–124, 2019, doi: 10.1007/s10462-018-09679-z.

[49] M. Openja, A. Nikanjam, A. H. Yahmed, F. Khomh, Z. Ming, and Jiang, "An Empirical Study of Challenges in Converting Deep Learning Models," Jun. 2022, [Online]. Available: http://arxiv.org/abs/2206.14322

[50] [50] K. Zhou, T. P. Michalak, and Y. Vorobeychik, "Adversarial robustness of similarity-based link prediction," Proceedings - IEEE International Conference on Data Mining, ICDM, vol. 2019-Novem, pp. 926–935, 2019, doi: 10.1109/ICDM.2019.00103.

[51] R. . ShyamChari, "IMAGE RECOGNITION AND STUDY OF HYPERPARAMETER OPTIMIZATION OF CONVOLUTIONAL NEURAL NETWORKS USING TENSORFLOW AND KERAS FRAMEWORKS Deep," PhD diss., 2018 Lyles College of Engineering California State University, Fresno May 2018, no. May 2018, p. 400, 2018.

[52] B. Qi, H. Sun, H. Zhang, and X. Gao, "Reusing Convolutional Neural Network Models through Modularization and Composition," Nov. 2023, [Online]. Available: http://arxiv.org/abs/2311.04438

[53] H. Kwon, P. Chatarasi, M. Pellauer, A. Parashar, V. Sarkar, and T. Krishna, "Understanding reuse, performance, and hardware cost of DNN dataflows: A data-centric approach," in Proceedings of the Annual International Symposium on Microarchitecture, MICRO, IEEE Computer Society, Oct. 2020, pp. 754–768. doi: 10.1145/3352460.3358252.

[54] J. C. Davis, P. Jajal, W. Jiang, T. R. Schorlemmer, N. Synovic, and G. K. Thiruvathukal, "Reusing Deep Learning Models: Challenges and Directions in Software Engineering," 2023.

[55] J. Ghofrani, E. Kozegar, A. Bozorgmehr, and M. D. Soorati, "Reusability in Artificial Neural Networks: An Empirical Study," in Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI), Gesellschaft fur Informatik (GI), 2021, pp. 37–38. doi: 10.18420/SE2020_07.

[56] M. Niazi, A. M. Saeed, M. Alshayeb, S. Mahmood, and S. Zafar, "A maturity model for secure requirements engineering," Comput Secur, vol. 95, Aug. 2020, doi: 10.1016/j.cose.2020.101852.

[57] C. W. Krueger, "Software Reuse."

[58] H. Hussain, P. S. Tamizharasan, and C. S. Rahul, "Design possibilities and challenges of DNN models: a review on the perspective of end devices," Artif Intell Rev, vol. 55, no. 7, pp. 5109–5167, Oct. 2022, doi: 10.1007/s10462-022-10138-z.

[59] F. Authors, F. Wu, T. Li, F. Luo, S. Wu, and C. Xiao, "A Survey on Deep Transfer Learning and Edge Computing for Mitigating the COVID-19 Pandemic," Journal of Systems Architecture, p. 32, 2020, doi: https://doi.org/10.1016/j.sysarc.2020.101830.

[60] X. Han et al., "Pre-trained models: Past, present and future," AI Open, vol. 2, pp. 225–250, Jan. 2021, doi: 10.1016/j.aiopen.2021.08.002.

[61] A. Rahman et al., "Machine learning and deep learning-based approach in smart healthcare: Recent advances, applications, challenges and opportunities," 2024, American Institute of Mathematical Sciences. doi: 10.3934/publichealth.2024004.

[62] H. V. Pham et al., "Problems and Opportunities in Training Deep Learning Software Systems: An Analysis of Variance," in Proceedings - 2020 35th IEEE/ACM International Conference on Automated Software Engineering, ASE 2020, Institute of Electrical and Electronics Engineers Inc., Sep. 2021, pp. 771–783. doi: 10.1145/3324884.3416545.

[63] P. Jajal et al., "Analysis of Failures and Risks in Deep Learning Model Converters: A Case Study in the ONNX Ecosystem," Mar. 2024, [Online]. Available: http://arxiv.org/abs/2303.17708

[64] V. Sze, Y. Chen, T. Yang, and J. Emer, "Efficient Processing of Deep Neural Networks: A Tutorial and Survey," Proceedings of the IEEE, vol. 105, no. 12, 2017, doi: 10.1109/JPROC.2017.2761740.

[65] S. Sengupta et al., "A review of deep learning with special emphasis on architectures, applications and recent trends," Knowl Based Syst, vol. 194, p. 105596, 2020, doi: 10.1016/j.knosys.2020.105596.

[66] Z. J. Yao, J. Bi, and Y. X. Chen, "Applying Deep Learning to Individual and Community Health Monitoring Data: A Survey," International Journal of Automation and Computing, vol. 15, no. 6, pp. 643–655, 2018, doi: 10.1007/s11633-018-1136-9.

[67] P. Giudici, M. Centurelli, and S. Turchetta, "Artificial Intelligence risk measurement," Expert Syst Appl, vol. 235, Jan. 2024, doi: 10.1016/j.eswa.2023.121220.

[68] N. Soni, E. K. Sharma, N. Singh, and A. Kapoor, "Artificial Intelligence in Business: From Research and Innovation to Market Deployment," Procedia Computer Science Journal / International Conference on Computational Intelligence and Data Science (ICCIDS 2019)., vol. 167, no. 2019, pp. 2200–2210, 2020, doi: 10.1016/j.procs.2020.03.272.

[69] A. Vinothini and S. B. Priya, "Survey of machine learning methods for big data applications," ICCIDS 2017 - International Conference on Computational Intelligence in Data Science, Proceedings, vol. 2018-Janua, pp. 1–5, 2018, doi: 10.1109/ICCIDS.2017.8272638.

[70] P. Singhal and S. Pareek, "Artificial neural network for prediction of breast cancer," Proceedings of the International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2018 IEEE Xplore, pp. 464–468, 2019, doi: 10.1109/I-SMAC.2018.8653700.

[71] G. A. Lewis, S. Bellomo, and A. Galyardt, "Component Mismatches Are a Critical Bottleneck to Fielding AI-Enabled Systems in the Public Sector," Carnegie Mellon Software Engineering Institute Pittsburgh, PA USA arXiv:1910.06136v1.AAAI, vol. 1, pp. 2–5, 2019.

[72] A. Sheth, M. Gaur, U. Kursuncu, and R. Wickramarachchi, "Shades of Knowledge-Infused Learning for Enhancing Deep Learning," IEEE Internet Comput, vol. 23, no. 6, pp. 54–63, 2019, doi: 10.1109/MIC.2019.2960071.

[73] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "A survey on deep learning for big data," Information Fusion, vol. 42, no. August 2017, pp. 146–157, 2018, doi: 10.1016/j.inffus.2017.10.006.

[74] A. Koutsoukas, K. Monaghan, X. Li, and J. Huan, "Deep-learning: Investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data," Journal of Cheminformatics Springer, vol. 9, no. 1, 2017, doi: 10.1186/s13321-017-0226-y.

[75] L. Liu et al., "Deep Learning for Generic Object Detection : A Survey," International Journal of Computer Vision (2020) 128:261–318 Springer, vol. 128, no. 2, pp. 261–318, 2021, doi: 10.1007/s11263-019-01247-4.

[76] A. Abrol, M. Bhattarai, A. Fedorov, Y. Du, S. Plis, and V. Calhoun, "Deep residual learning for neuroimaging: An application to predict progression to Alzheimer's disease," J Neurosci Methods, vol. 339, no. September 2019, p. 108701, 2021, doi: 10.1016/j.jneumeth.2020.108701.

[77] M. A. Ebrahimighahnavieh, S. Luo, and R. Chiong, "Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review," Comput Methods Programs Biomed, vol. 187, p. 105242, 2021, doi: 10.1016/j.cmpb.2019.105242.

[78] N. Rahim, S. El-Sappagh, S. Ali, K. Muhammad, J. Del Ser, and T. Abuhmed, "Prediction of Alzheimer's progression based on multimodal Deep-Learning-based fusion and visual Explainability of time-series data," Information Fusion, vol. 92, no. June 2022, pp. 363–388, 2023, doi: 10.1016/j.inffus.2022.11.028.

[79] D. van Tilborg, H. Brinkmann, E. Criscuolo, L. Rossen, R. Özçelik, and F. Grisoni, "Deep learning for low-data drug discovery: Hurdles and opportunities," Jun. 01, 2024, Elsevier Ltd. doi: 10.1016/j.sbi.2024.102818.

[80] Y. Liu, C. Chen, R. Zhang, M. Research, H. Lin, and M. Yang, "MMdnn: Enhancing the Interoperability between Deep Learning Frameworks by

Model Conversion," pp. 6–8, 2020, [Online]. Available: https://en.wikipedia.org/wiki/Interoperability.

[81] C. Jiang, Z. Huang, T. Pedapati, P. Y. Chen, Y. Sun, and J. Gao, "Network properties determine neural network performance," Nat Commun, vol. 15, no. 1, Dec. 2024, doi: 10.1038/s41467-024-48069-8.

[82] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "A Survey of Model Compression and Acceleration for Deep Neural Networks," IEEE SIGNAL PROCESSING MAGAZINE, SPECIAL ISSUE ON DEEP LEARNING FOR IMAGE UNDERSTANDING, no. SPECIAL ISSUE, Oct. 2020, [Online]. Available: http://arxiv.org/abs/1710.09282

[83] F. Zhuang et al., "A Comprehensive Survey on Transfer Learning," Nov. 2021, [Online]. Available: http://arxiv.org/abs/1911.02685

[84] L. Liu, Y. Wu, W. Wei, W. Cao, S. Sahin, and Q. Zhang, "Benchmarking deep learning frameworks: Design considerations, metrics and beyond," Proceedings - International Conference on Distributed Computing Systems; IEEE Computer Society, vol. 2018-July, pp. 1258–1269, 2018, doi: 10.1109/ICDCS.2018.00125.

[85] S. S. Khanal, P. W. C. Prasad, A. Alsadoon, and A. Maag, "A systematic review: machine learning based recommendation systems for e-learning," Educ Inf Technol (Dordr), vol. 25, no. 4, pp. 2635–2664, Jul. 2020, doi: 10.1007/s10639-019-10063-9.

[86] M. Perano, G. L. Casali, Y. Liu, and T. Abbate, "Professional reviews as service: A mix method approach to assess the value of recommender systems in the entertainment industry," Technol Forecast Soc Change, vol. 169, no. April, p. 120800, 2021, doi: 10.1016/j.techfore.2021.120800.

[87] B. Zoph and J. Shlens, "Learning Transferable Architectures for Scalable Image Recognition," 2018.

[88] J. Li et al., "Networked human motion capture system based on quaternion navigation," in BodyNets International Conference on Body Area Networks, 2018. doi: 10.1145/0000000.0000000.

[89] A. Goyal et al., "Automatic border surveillance using machine learning in remote video surveillance systems," in Lecture Notes in Electrical Engineering, vol. 569, Springer Verlag, 2020, pp. 751–760. doi: 10.1007/978-981-13-8942-9_64.

[90] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," Comput Struct Biotechnol J, vol. 13, pp. 8–17, 2017, doi: 10.1016/j.csbj.2014.11.005.

[91] et al Noureen Talpur, "Deep Neuro Fuzzy System application trends, challenges, and future perspectives; A systematic Review.pdf," 2023.

[92] O. Ali, W. Abdelbaki, A. Shrestha, E. Elbasi, M. A. A. Alryalat, and Y. K. Dwivedi, "A systematic literature review of artificial intelligence in the healthcare sector: Benefits, challenges, methodologies, and functionalities," Journal of Innovation and Knowledge, vol. 8, no. 1, 2023, doi: 10.1016/j.jik.2023.100333.

[93] A. Darwish, A. E. Hassanien, and S. Das, "A survey of swarm and evolutionary computing approaches for deep learning," Artif Intell Rev, vol. 53, no. 3, 2020, doi: 10.1007/s10462-019-09719-2.

[94] G. A. Lewis, S. Bellomo, and I. Ozkaya, "Characterizing and Detecting Mismatch in Machine-Learning-Enabled Systems," Mar. 2021, [Online]. Available: http://arxiv.org/abs/2103.14101

[95] J. R. Ragini, P. M. R. Anand, and V. Bhaskar, "Big data analytics for disaster response and recovery through sentiment analysis," Int J Inf Manage, vol. 42, no. May, pp. 13–24, 2018, doi: 10.1016/j.ijinfomgt.2018.05.004.

[96] B. Kumar Padhi, S. S. Nayak, and B. N. Biswal, "Machine Learning for Big Data Processing: A Literature Review," INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN TECHNOLOGY; IJIRT 140001 December 2018 | IJIRT | Volume 5 Issue 7 | ISSN: 2349-6002 samples, vol. 5, no. 7, p. 10, 2018, [Online]. Available: https://www.researchgate.net/publication/331497335

[97] D. A. Neu, J. Lahann, and P. Fettke, "A systematic literature review on state-of-the-art deep learning methods for process prediction," 2022. doi: 10.1007/s10462-021-09960-8.

[98] B. Krishna, N. Rekulapelli, and B. P. Kauda, "Comparison of different deep learning frameworks," Mater Today Proc, no. xxxx, p. 5, 2020, doi: 10.1016/j.matpr.2020.09.608.

[99] A. Darwish, A. Ella, and H. Swagatam, "A survey of swarm and evolutionary computing approaches for deep learning," Artif Intell Rev, 2020, doi: 10.1007/s10462-019-09719-2.

[100] M. Modasshir, A. Quattrini Li, and I. Rekleitis, "Deep neural networks: A comparison on different computing platforms," Proceedings - 2018 15th Conference on Computer and Robot Vision, CRV 2018 978-1-5386-6481-0/18/$31.00 ©2018 IEEE, pp. 383–389, 2018, doi: 10.1109/CRV.2018.00060.

[101] J. Dai and A. Gibson, "Comparison of deep-learning software originally," https://en.wikipedia.org/wiki/Comparison_of_deep-learning_software, pp. 1–5, 2020.

[102] Y. Liu, C. Chen, R. Zhang, M. Research, H. Lin, and M. Yang, "Enhancing the Interoperability between Deep Learning Frameworks by Model Conversion Tingting Qin," 2020.

[103] S. Hassan, "Deep context of citations using machine learning models.pdf," 2018.

[104] Khronos G, "Neural Network Exchange Format (NNEF)," 2019.

[105] W. F. Lin et al., "ONNC: A Compilation Framework Connecting ONNX to Proprietary Deep Learning Accelerators," Proceedings 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems, AICAS 2019, pp. 214–218, 2019, doi: 10.1109/AICAS.2019.8771510.

[106] S. Sanchez-Gordon and S. Luján-Mora, "Technological Innovations in Large-Scale Teaching: Five Roots of Massive Open Online Courses," Journal of Educational Computing Research, vol. 56, no. 5, pp. 623–644, 2018, doi: 10.1177/0735633117727597.

[107] H. N. F, Y. S. S, and S. S. M, "A Systematic Literature Review on Features of Deep Learning in Big Data Analytics," Int. J. Advance Soft Compu. Appl, vol. 9, no. 1, 2017.

[108] F. Hujainah, R. B. A. Bakar, M. A. Abdulgabber, and K. Z. Zamli, "Software requirements prioritisation: a systematic literature review on significance, stakeholders, techniques and challenges," IEEE Access, vol. 6, pp. 71497–71523, 2018.

[109] L. A. Kahale et al., "PRISMA flow diagrams for living systematic reviews: a methodological survey and a proposal," F1000Res, vol. 10, p. 192, Mar. 2022, doi: 10.12688/f1000research.51723.1.

[110] Kitchenham, "Guidelines for performing Systematic Literature Reviews in Software Engineering," p. 65, 2007, doi: 10.1145/1134285.1134500.

[111] S. Mittal and S. Vaishay, "A survey of techniques for optimizing deep learning on GPUs," Journal of Systems Architecture; Publisher: ELSEVIER , RADARWEG 29, AMSTERDAM, NETHERLANDS, 1043 NX, vol. 99, p. 45, 2019, doi: 10.1016/j.sysarc.2019.101635.

[112] S. Fatima, K. C. Desouza, and G. S. Dawson, "National strategic artificial intelligence plans: A multi-dimensional analysis," Journal of Economic Analysis and Policy; Publisher: ELSEVIER , RADARWEG 29, AMSTERDAM, NETHERLANDS, 1043 NX, vol. 67, pp. 178–194, 2020, doi: 10.1016/j.eap.2020.07.008.

[113] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," IEEE Communications Surveys and Tutorials, vol. 20, no. 4, pp. 2595–2621, 2018, doi: 10.1109/COMST.2018.2846401.

[114] A. Lavecchia, "Deep learning in drug discovery: opportunities, challenges and future prospects," Drug Discovery Today Journal 1359-6446/ã 2019 Elsevier Ltd. All rights reserved SCI, vol. 24, no. 10, p. 16, 2019, doi: 10.1016/j.drudis.2019.07.006.

[115] T. R. Rao, P. Mitra, R. Bhatt, and A. Goswami, "The big data system, components, tools, and technologies: a survey," Knowledge and Information Systems ISI : Science Citation Index Expanded, vol. 60, no. 3, pp. 1165–1245, 2019, doi: 10.1007/s10115-018-1248-0.

[116] Y. Zahidi, Y. El Younoussi, and Y. Al-amrani, "A powerful comparison of deep learning frameworks for Arabic sentiment analysis," International Journal of Electrical and Computer Engineering (IJECE), vol. 11, no. 1, pp. 745–752, 2021, doi: 10.11591/ijece.v11i1.pp745-752.

[117] M. Salvi, U. R. Acharya, F. Molinari, and K. M. Meiburger, "The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis," Comput Biol Med, vol. 128, p. 104129, 2021, doi: 10.1016/j.compbiomed.2020.104129.

[118]A. Shatnawi, G. Al-Bdour, R. Al-Qurran, and M. Al-Ayyoub, "A comparative study of open source deep learning frameworks," 2018 9th International Conference on Information and Communication Systems, ICICS 2018; IEEE, vol. 2018-Janua, no. April, pp. 72–77, 2018, doi: 10.1109/IACS.2018.8355444.

[119]A. Oussous, F. Z. Benjelloun, A. Ait Lahcen, and S. Belfkih, "Big Data technologies: A survey," Journal of King Saud University - Computer and Information Sciences ISI: Science Citation Index Expanded, vol. 30, no. 4, pp. 431–448, 2018, doi: 10.1016/j.jksuci.2017.06.001.

[120]N. Mahendran, P. M. D. R. Vincent, and K. Srinivasan, "Machine Learning Based Computational Gene Selection Models : A Survey , Performance Evaluation , Open Issues , and Future Research Directions," Frontiers in genetics Journal SCIE (web of science), vol. 11, no. December, pp. 1–25, 2020, doi: 10.3389/fgene.2020.603808.

[121]R. Vinayakumar, K. Soman, and P. Poornachandran, "Evaluation of recurrent neural network and its variants for intrusion detection system (IDS)," International Journal of Information System Modeling and Design, vol. 8, no. 3, 2017, doi: 10.4018/IJISMD.2017070103.

[122]A. Awan, K. Manian, C. Chu, H. Subramoni, and D. Panda, "Optimized large-message broadcast for deep learning workloads: MPI, MPI+NCCL, or NCCL2?," Parallel Comput, vol. 85, 2019, doi: 10.1016/j.parco.2019.03.005.

[123]D. W. Bates, A. Auerbach, P. Schulam, A. Wright, and S. Saria, "Reporting and Implementing Interventions Involving Machine Learning and Artificial Intelligence," Annals of internal medicine Jour, vol. 172, no. 11, 2020, doi: 10.7326/M19-0872.

[124]W. Dai and D. Berleant, "Benchmarking contemporary deep learning hardware and frameworks: A survey of qualitative metrics," Proceedings - 2019 IEEE 1st International Conference on Cognitive Machine Intelligence, CogMI 2019, pp. 148–155, 2019, doi: 10.1109/CogMI48466.2019.00029.

[125]Y. Chen, Y. Tian, and M. He, "Monocular human pose estimation: A survey of deep learning-based methods," Computer Vision and Image Understanding, vol. 192, no. December 2019, p. 102897, 2020, doi: 10.1016/j.cviu.2019.102897.

[126]X. Lu, H. Shi, R. Biswas, M. Javed, and D. Panda, "DLoBD: A Comprehensive Study of Deep Learning over Big Data Stacks on HPC Clusters," IEEE Transactions on Multi-Scale Computing Systems, vol. 4, no. 4, 2018, doi: 10.1109/TMSCS.2018.2845886.

[127]U. Naseem, I. Razzak, and P. W. Eklund, "A survey of pre-processing techniques to improve short-text quality: a case study on hate speech detection on twitter," Multimedia Tools and Applications Journal; Publisher: SPRINGER , VAN GODEWIJCKSTRAAT 30, DORDRECHT, NETHERLANDS, 3311 GZ, p. 28, 2020, doi: 10.1007/s11042-020-10082-6.

[128]A. Jain, A. A. Awan, H. Subramoni, and D. K. Panda, "Scaling TensorFlow, PyTorch, and MXNet using MVAPICH2 for high-performance deep learning on Frontera," Proceedings of DLS 2019: Deep Learning on Supercomputers - Held in conjunction with SC 2019: The International Conference for High Performance Computing, Networking, Storage and Analysis, 2019, doi: 10.1109/DLS49591.2019.00015.

[129]T. Meng, X. Jing, Z. Yan, and W. Pedrycz, "A survey on machine learning for data fusion," Information Fusion, vol. 57, no. 2, pp. 115–129, 2020, doi: 10.1016/j.inffus.2019.12.001.

[130]F. Zaki, A. E. Mohamed, and S. G. Sayed, "CtuNet : A Deep Learning-based Framework for Fast CTU Partitioning of H265 / HEVC Intra-coding," Ain Shams Engineering Journal, no. xxxx, pp. 1–8, 2021, doi: 10.1016/j.asej.2021.01.001.

[131]M. Li et al., "The Deep Learning Compiler: A Comprehensive Survey," IEEE Transactions on Parallel and Distributed Systems; Publisher: IEEE COMPUTER SOC , 10662 LOS VAQUEROS CIRCLE, PO BOX 3014, LOS ALAMITOS, USA, CA, 90720-1314, vol. 32, no. 3, pp. 708–727, 2021, doi: 10.1109/TPDS.2020.3030548.

[132]Y. Oyama, S. Matsuoka, T. Ben-Nun, and T. Hoefler, "Accelerating Deep Learning Frameworks with Micro-Batches," Proceedings - IEEE International Conference on Cluster Computing, ICCC, vol. 2018-, no. 11, 2018, doi: 10.1109/CLUSTER.2018.00058.

[133]T. Doleck, D. J. Lemay, R. B. Basnet, and P. Bazelais, "Predictive analytics in education: a comparison of deep learning frameworks," Educ Inf Technol (Dordr), vol. 25, no. 3, pp. 1951–1963, 2020, doi: 10.1007/s10639-019-10068-4.

[134]E. M. Dogo, N. I. Nwulu, B. Twala, and C. Aigbavboa, "A survey of machine learning methods applied to anomaly detection on drinking-water quality data," Urban Water J, vol. 16, no. 3, pp. 235–248, 2019, doi: 10.1080/1573062X.2019.1637002.

[135]J. H. Tao et al., "BenchIP: Benchmarking Intelligence Processors," J Comput Sci Technol, vol. 33, no. 1, 2018, doi: 10.1007/s11390-018-1805-8.

[136]C. L. Hung, Y. Y. Lin, C. Y. Tang, C. Wang, and M. C. Chen, "Performance of convolution neural network based on multiple GPUs with different data communication models," Proceedings - 2018 IEEE/ACIS 19th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2018, pp. 87–92, 2018, doi: 10.1109/SNPD.2018.8441056.

[137]L. Zhou, S. Pan, J. Wang, and A. V Vasilakos, "Neurocomputing Machine learning on big data : Opportunities and challenges," ELSEVIER JOURNAL Science Direct, vol. 237, no. September 2016, pp. 350–361, 2017, doi: 10.1016/j.neucom.2017.01.026.

[138]S. Mao, B. Wang, Y. Tang, and F. Qian, "Opportunities and Challenges of Artificial Intelligence for Green Manufacturing in the Process Industry," Engineering Journal; Publisher: ELSEVIER , RADARWEG 29, AMSTERDAM, NETHERLANDS, 1043 NX, vol. 5, no. 6, pp. 995–1002, 2019, doi: 10.1016/j.eng.2019.08.013.

[139]L. E. Lwakatare, A. Raj, I. Crnkovic, J. Bosch, and H. H. Olsson, "Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions," Information and Software Technology 127 (2020) 106368 Contents ELSEVIER, vol. 127, no. June, p. 17, 2020, doi: 10.1016/j.infsof.2020.106368.

[140]E. Breck, N. Polyzotis, S. Roy, S. E. Whang, and M. Zinkevich, "DATA VALIDATION FOR MACHINE LEARNING," Proceedings of the 2 nd SysML Conference, Palo Alto, CA, USA, 2019, pp. 1--14, 2019.

[141]Y. Roh, G. Heo, and S. E. Whang, "A survey on data collection for machine learning: A big data - AI integration perspective," IEEE Transactions on Knowledge and Data Engineering. Publisher: IEEE COMPUTER SOC , 10662 LOS VAQUEROS CIRCLE, PO BOX 3014, LOS ALAMITOS, USA, CA, 90720-1314, vol. 2, pp. 1–20, 2018, doi: 10.1109/tkde.2019.2946162.

[142]G. Al-bdour, R. Al-qurran, M. Al-ayyoub, A. Shatnawi, and M. Al-ayyoub, "Benchmarking open source deep learning frameworks," International Journal of Electrical and Computer Engineering (IJECE), vol. 10, no. 5, pp. 5479–5486, 2020, doi: 10.11591/ijece.v10i5.pp5479-5486.

[143]P. Xu, S. Shi, and X. Chu, "Performance Evaluation of Deep Learning Tools in Docker Containers," Proceedings - 2017 3rd International Conference on Big Data Computing and Communications, BigCom 2017, pp. 395–403, 2017, doi: 10.1109/BIGCOM.2017.32.

[144]M. Štufi and B. Baˇ, "Big Data Analytics and Processing Platform in Czech Republic Healthcare," Applied Science Journal . Appl. Sci. 2020, 10, 1705; doi:10.3390/app10051705 www.mdpi.com/journal/applsci10, 1705. MDPI, no. Appl. Sci. 2020, 10, 1705; doi:10.3390/app10051705, p. 23, 2020.

[145]F. Cappa, R. Oriani, E. Peruffo, and I. McCarthy, "Big Data for Creating and Capturing Value in the Digitalized Environment: Unpacking the Effects of Volume, Variety, and Veracity on Firm Performance*," Journal of Product Innovation Management Publisher: WILEY , 111 RIVER ST, HOBOKEN, USA, NJ, 07030-5774 ISI: Science Citation Index Expanded | Social Sciences Citation Index, vol. 0, no. 0, pp. 1–19, 2020, doi: 10.1111/jpim.12545.

[146]E. Kusmenko, S. Nickels, S. Pavlitskaya, B. Rumpe, and T. Timmermanns, "Modeling and Training of Neural Processing Systems," In: Conference on Model Driven Engineering Languages and Systems (MODELS'19), pp. 283--293, IEEE, Munich, Sep. 2019. www.se-rwth.de/publications/, p. 11, 2019.

[147]L. F. Sánchez-Peralta, L. Bote-Curiel, A. Picón, F. M. Sánchez-Margallo, and J. B. Pagador, "Deep learning to find colorectal polyps in colonoscopy: A systematic literature review," Artificial Intelligence in Medicine; Publisher: ELSEVIER , RADARWEG 29, AMSTERDAM, NETHERLANDS, 1043 NX, vol. 108, no. March, p. 101923, 2020, doi: 10.1016/j.artmed.2020.101923.

[148]S. Bianco, R. Cadène, L. Celona, and P. Napoletano, "Benchmark Analysis of Representative Deep Neural Network Architectures," IEEE Access IEEE TRANSACTIONS and JOURNALS, vol. 6, pp. 64270–64277, 2018, doi: 10.1109/ACCESS.2018.2877890.

[149]Q. Guo et al., "An Empirical Study towards Characterizing Deep Learning Development and Deployment across Different Frameworks and Platforms," Sep. 2019, [Online]. Available: http://arxiv.org/abs/1909.06727

[150]L. Messeri and M. J. Crockett, "Artificial intelligence and illusions of understanding in scientific research," Nature, vol. 627, no. 8002, pp. 49–58, Mar. 2024, doi: 10.1038/s41586-024-07146-0.

[151]X. Li et al., "Interpretable Deep Learning: Interpretation, Interpretability, Trustworthiness, and Beyond," Mar. 2021, [Online]. Available: http://arxiv.org/abs/2103.10689

[152]J. J. Pfeiffer, J. Neville, and P. N. Bennett, "Overcoming relational learning biases to accurately predict preferences in large scale networks," WWW 2015 - Proceedings of the 24th International Conference on World Wide Web; ACM 978-1-4503-3469-3/15/05, pp. 853–863, 2015, doi: 10.1145/2736277.2741668.

[153]D. Jeong, "Artificial Intelligence Security Threat, Crime, and Forensics: Taxonomy and Open Issues," IEEE Access, vol. 8, 2020, doi: 10.1109/access.2020.3029280.

[154]A. Mendelson, "Security and Privacy in the Age of Big Data and Machine Learning," Computer Journal; Publisher: IEEE COMPUTER SOC , 10662 LOS VAQUEROS CIRCLE, PO BOX 3014, LOS ALAMITOS, USA, CA, 90720-1314 IEEE, vol. 52, pp. 65–70, 2019, doi: 10.1109/MC.2019.2943137.

[155]C. Issues, A. Projects, and W. Security, "Caffe to TensorFlow," 2017, [Online]. Available: https://github.com/ethereon/caffe-tensorflow

[156]H. Park, J. H. Lee, Y. Oh, S. Ha, and S. Lee, "Privacy attacks against deep learning models and their countermeasures," Journal of Systems Architecture journal, vol. 9, no. 114 (2021) 101940, 2021, [Online]. Available: www.elsevier.com/locate/sysarc%0APrivacy

[157]C. Ning and F. You, "Optimization under Uncertainty in the Era of Big Data and Deep Learning : When Machine Learning Meets Mathematical Programming," Computers & Chemical Engineering ; Publisher: PERGAMON-ELSEVIER SCIENCE LTD , THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD, ENGLAND, OX5 1GB, no. 607, pp. 1–42, 2020.

[158]L. Liu, Y. Wu, W. Wei, W. Cao, S. Sahin, and Q. Zhang, "Benchmarking Deep Learning Frameworks : Design Considerations , Metrics and Beyond," 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS) IEEE computer Society, no. July 2019, pp. 1258–1269, 2018, doi: 10.1109/ICDCS.2018.00125.

[159]P. Mattson et al., "MLPerf: An Industry Standard Benchmark Suite for Machine Learning Performance," IEEE MICRO, Published by the IEEE Computer Society, no. 0272-1732 2020 IEEE, p. 9, 2020.

[160]Q. Zhang et al., "A Survey on Deep Learning Benchmarks: Do We Still Need New Ones?," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer, 2019, pp. 36–49. doi: 10.1007/978-3-030-32813-9_5.

[161]P. Jajal et al., "Analysis of Failures and Risks in Deep Learning Model Converters: A Case Study in the ONNX Ecosystem," Mar. 2023, [Online]. Available: http://arxiv.org/abs/2303.17708

[162]A. Paszke et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library," 2020.

[163]S. Ghemawat et al., "TensorFlow: A system for large-scale machine learning," Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2016, p. 21, 2016.

[164]T. Chen et al., Open access to the Proceedings of the 13th USENIX Symposium on Operating Systems Design and Implementation is sponsored by USENIX. TVM: An Automated End-to-End Optimizing Compiler for Deep Learning TVM: An Automated End-to-End Optimizing Compiler for Deep Learning. 2018. [Online]. Available: https://www.usenix.org/conference/osdi18/presentation/chen

[165]I. Soliman, "Build ONNX Runtime from Source on Windows 10," pp. 1–15, 2020.

A Machine Learning Operations (MLOps) Monitoring Model Using BI-LSTM and SARSA Algorithms

Zeinab Shoieb Elgamal, Laila Elfangary, Hanan Fahmy
Department of Information Systems-Faculty of Computers and Artificial Intelligence, Helwan University, Helwan, Egypt[1]

***Abstract*—Machine learning operations (MLOps) achieves faster model development, deliver higher machine learning models quality, and faster deployment cycle. Unfortunately, MLOps is still an uncertain concept with ambiguous research implications. Professionals and academics have focused only on creating machine learning models, rather than using sophisticated machine learning systems in practical situations. Furthermore, the monitoring system must have a comprehensive view over the system interactions. The need for a strong efficient monitoring system increases when it comes to use the multi container services. Therefore, this research provides a new proposed model called Multi Containers Monitoring (MCM) Model, based on multi container service and machine learning approaches which are bidirectional long short-term memory (BI-LSTM) and state-action-reward-state-action (SARSA). The proposed MCM model enables MLOps systems to be scaled and monitored efficiently. The proposed MCM model realizes and interprets the interactions between the containers. The proposed MCM model enhances the performance of the software release and increases the number of software deployments across different types of environments. Moreover, this research proposes four routines for each layer of the proposed MCM model that illustrates how each layer is going to be developed. This research also illustrates how the proposed MCM model achieves improvements ratio in software deployment cycles by using MLOps up to 24.55% and in build duration cycle up to 13%.**

Keywords*—Machine learning; MLOps; monitoring; container; model

I. Introduction

The fast and increasing popularity of machine learning (ML) applications has led to growing attention in Machine Learning Operations (MLOps), that is, the practice of continuous integration and deployment (CI/CD) of ML-enabled systems [1]. Since changes may not affect only the code but also the ML model parameters and the data themselves, the automation of traditional CI/CD needs to be extended to outspread to monitor model retraining in production [1]. ML has become a significant technique to leverage the potential of data and allows businesses to be more innovative, efficient, and sustainable [2] [3] [4] [5]. However, the success of many ML applications in the real world doesn't meet expectations as the ML community has focused extensively on the building of ML models not on building production-ready ML products and providing the necessary coordination of the resulting [2] [6] [7] [8]. Besides that, these applications started to produce and maintain a huge amount of data from their operations. Those new developments require monitoring the operations of applications in real-time [9]. If MLOps model selection and training are not closely and carefully monitored, applications may lose value in the market and organizations might be at risk of losing money, but the worst is to lose their reputation [10] [11] [12].

This research proposes a multi-container monitoring (MCM) model to monitor the communication and all containers' behavior for the software deployment cycles to help in more frequent releases and reduce production issues. Further, discusses the MLOps practice to effectively handle the issue of creating and monitoring effective ML. Furthermore, adopts a broad viewpoint to provide comprehension of the relevant principles, responsibilities, and architectural structures.

This research makes a significant contribution to the software industry by:

- Review all the previous studies in monitoring containerized software.
- Delineate what kind of problems MLOps practices may be best suited to apply, to help in reducing the re-developing and re-deploying.
- Monitor software performance at a finer granularity level.
- Monitor DevOps and MLOps pipelines and system infrastructure behavior.
- Reduce the build time of software systems.
- Improve the deployment rates of existing methods in software systems.

This research applies a new machine learning technique to monitor and learn more ML model features based on different software systems.

The remainder of this research is structured as follows. Section II presents the background. Section III illustrates the necessary definitions and related work in the field. Section IV, presents an overview of the utilized methodology and the proposed MCM model. It also presents the MCM model challenges and limitations. Section V and VI presents model dataset and model configuration. Results and discussion is

given in Section VII and Section VIII respectively. Finally, Section VI, concludes the work with a summary.

II. BACKGROUND

There are a variety of software process models and development methodologies used in software engineering such as waterfall and the agile manifesto. Those methodologies have similar objectives, which are to deliver production-ready software applications [2] [13] [14]. Recently software development teams have moved away from the traditional waterfall methodology to DevOps as the traditional life cycle is not suited for dynamic projects as needed in the ML development process, as shown in Fig. 1.

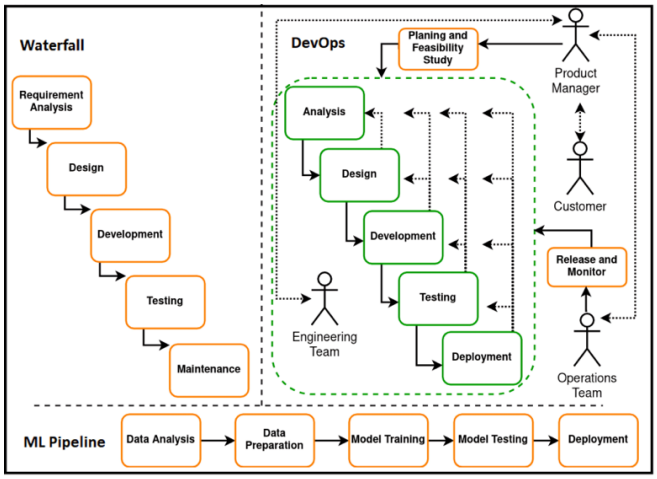


Fig. 1. Waterfall, DevOps SDLC, and the manual ML pipeline [15].

DevOps refers to the modern software deployment model that combines software development (Dev) and IT operations (Ops) [16]. DevOps aims to enable automation, continuous integration, continuous deployment, monitoring, and team collaboration of software applications in fast and small releases [17] [18]. The two primary DevOps practices are Continuous Integration and Continuous Delivery.

Continuous integration (CI) is a software practice that concentrates on automating the creation and integration of code from many developers. To enable quicker development cycles and enhance quality, developers are demanded to merge their code into the primary repository more frequently in this procedure. Version control systems (VCS), automated software development, and testing procedures are the key elements of this practice [15] [19].

Continuous delivery (CD) core purpose is to deliver newly created features to the end user as rapidly as possible by building the software in a way that is constantly in a production-ready state to ensure that code updates might be released on demand fast and safely [15] [20] [21].

Continuous deployment (CDE), which is frequently confused with CD, is a different technique. Continuous deployment is a technique where every software modification is automatically pushed to production. Even so, some businesses have procedures in place for obtaining outside approval before releasing new application's version to users. Thus, continuous delivery is considered necessary in certain circumstances; however, continuous deployment is optional and can be skipped [15] [19] as shown in Fig. 2.

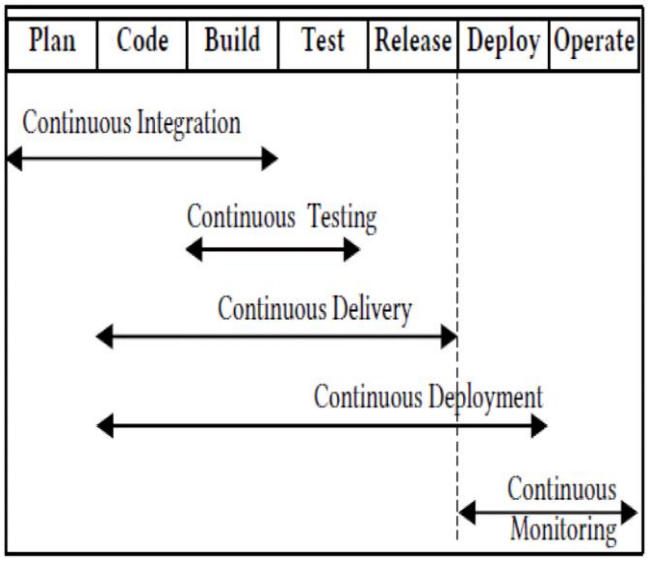


Fig. 2. Deployment pipelines [18].

Continuous monitoring is an automated procedure that uses cloud services to assess a deployed application's operational functionality against business criteria as it is being used [18] [22].

The DevOps pipeline, also known as the CI/CD pipeline, enables greater support for the deployment of applications to the cloud and utilizing a wide range of tools [15] [17] [23].

ML pipeline is defined as an automation of the ML life cycle by minimizing human interaction in routine processes [24].

MLOps refers to the complete vision of best practices and procedures from the design of the training data through the final deployment lifecycle [16] as shown in Fig. 3. MLOps can alternatively be considered as the integration of DevOps with machine learning techniques [25]. By another word, MLOps is the artificial intelligence (AI) equivalent of DevOps [16]. Furthermore, MLOps places a strong emphasis on automation while monitoring each step of the machine learning process, much like DevOps [26].

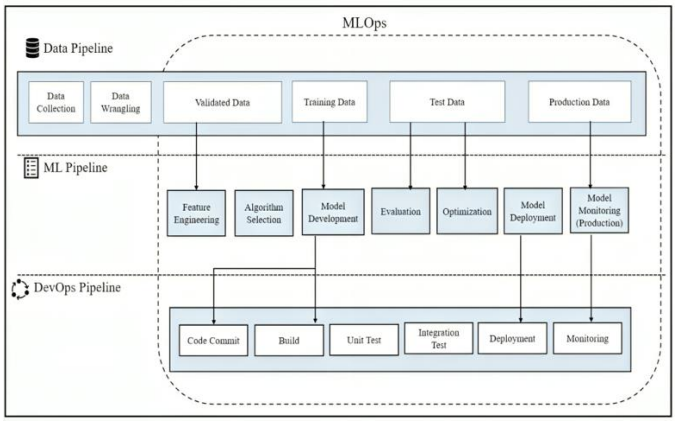


Fig. 3. High-level process perspective of MLOps [24].

MLOps phases are typically related to various roles and concepts, such as containerization and dockerizing [24] [27]. Containerization has become a standard approach for environments, which support on-demand, short-lived execution of computational tasks such as Function-as-a-Service (FaaS) platforms [28] [29].

Running applications in containers enables automatic orchestration and agile DevOps practices, in cloud-native platforms, the design of container and objects in software object-oriented programming (OOP) is similar: each container has a particular duty to carry out effectively [30] [31]. Cloud-native containers are easier to grow horizontally and replace, reuse, and update transparently [31], this has been termed the "Single Concern Principle" [32].

Dockerizing facilitates the hosting and execution of any kind of software applications, platforms, middleware, databases, packaged, in-house, and custom-built software. Moreover, the quicker maturation and reliability of the Docker platform have made it much easier to develop, distribute, deliver, and deploy software [33]. Furthermore, dockerizing offers a simple approach to isolate the network and restrict how much resource the containers can use [34].

There is another design concept that helps in monitoring the ML projects if it had been used during the project development phase which is the "Microservices" design rather than the monolithic design. Microservice architecture (MSA) is suggested to divide single-component applications into numerous loosely linked and independent microservice components [35] [36]. Microservices are applications broken down into their core functionalities. Each function operates as its own "service" inside a container and interacts with other containers across the network [37]. Microservices have several benefits over monolithic applications, such as autonomous update cycles, fine-grained resource control, and high elasticity [37]. The monolithic design is only appropriate for small-scale systems with straightforward internal structures since these monolithic applications adhere to an all-in-one architecture in which all functional modules are created and configured into precisely one deployment unit, namely, one container [38].

III. Related Work

Raúl Minin, et al., introduced a tool named Pangea that generates adequate execution settings for deploying analytic pipelines automatically. These pipelines are broken down into several stages so that each can be executed in the edge, fog, cloud, or on-premises environment which will minimize latency and make the best use of available hardware and software resources. Pangea is focused on achieving three specific goals: (1) creating the required infrastructure if it doesn't already exist; (2) providing it with the components needed to run the pipelines (i.e., configuring each host operating system and software, installing dependencies, and downloading the executable code); and (3) deploying the pipelines [39]. Raúl introduced a complex tool that takes a lot of work to conceptualize and build. Although the first version of the tool is sufficiently developed to demonstrate some of its potential advantages, further use cases and technology and connection compatibility must be added before it can be used in more situations. The web client requires to be improved to assist the management of users, pipelines, and infrastructure since Pangea isn't built to support the description and deployment of analytical pipelines in the training stage. Moreover, Pangea doesn't support monitoring pipelines and infrastructure behavior.

Matteo Testi, et al., provided a literature review on MLOps to illustrate the present difficulties in developing and sustaining an ML system in a production context. The literature review revealed that the utilization of MLOps in the workplace and the application of DevOps principles to machine learning are still under-discussed issues in academia. Furthermore, organizations will need to conduct experimental work to test the ML pipeline as they attempt to apply an ML approach to an end-to-end use case, going through each step and demonstrating what results if certain phases are skipped [40].

Sergio Moreschini, et al., offered a better illustration of MLOps by integrating ML development stages into the established DevOps practices. The research suggested a MLOps pipeline that concentrated on the duality between software engineers and machine learning developers and their roles [41]. Sergio's roadmap increased adoption of ML-based software generated a demand for ML developers who need to perform tasks in parallel to software developers and produced two extra loops for both the ML and software sides.

Pinchen Cui focused on providing security for containerization through secure monitoring of containerized applications to give better simulation of actual application behaviors and greater coverage of attacks with extended feature space [35]. Pinchen's research has not been put through an online evaluation. The elements of the security monitoring target must be enhanced to allow the framework to automatically determine what to monitor in an unsupervised manner. Meanwhile, the framework did not support scaling the dataset with various application architectures, such as multi-container applications, where a service is composed of several containers, and Docker Swarm's distributed monitoring.

Holger Gantikow, et al., suggested integrating containerized environments with rule-based security monitoring. The suitability of the method is investigated for both (1) a variety of undesirable behaviors that may point to abuse and attacks of workloads running inside a container and (2) misconfigurations and attempts to increase privileges and weaken isolation safeguards at the container runtime level [36]. The article does not cover the security monitoring of distributed workloads because shared workloads interact strongly across host borders.

While recent studies cover a variety of specific MLOps topics, a comprehensive conception, generalization, and explanation of ML systems monitor are still lacking. Different interpretations of the phrase "MLOps" may result in misconceptions, which may result in setup errors for the entire ML system. MCM model will work on the monitoring of the development for the multi-container "distributed" ML systems resulting in the ability to improve the software build and deployment cycles in real-world settings.

IV. THE MCM PROPOSED MODEL

The proposed Multi Container Monitoring (MCM) Model consists of four different layers as shown in Fig. 4. The MCM's model layers are the "Development layer, MLOps and container layer, and Monitoring layer with the support of different tools in the Tools and Automation layer".

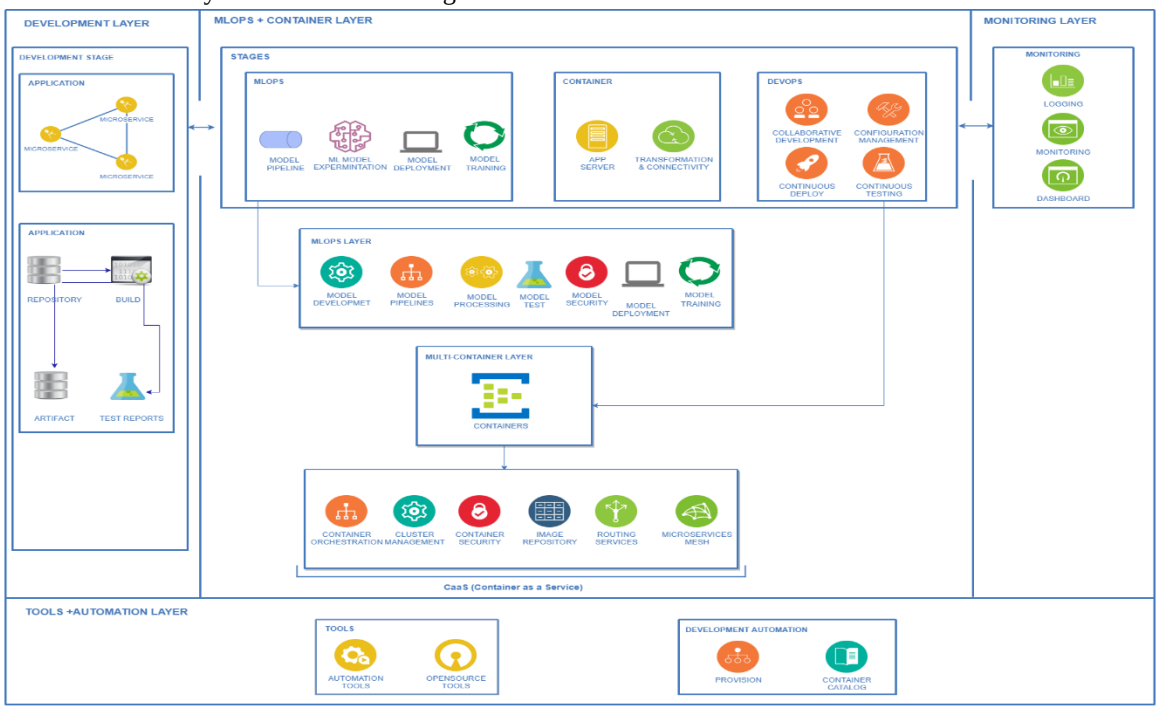


Fig. 4. Multi Containers Monitoring (MCM) model.

The MCM Model starts from the developer's commitment of a specific piece of code and ends with its deployment and monitoring on various environments. The next sub-sections will introduce a detailed description of the MCM's model layers and describe the components of the MCM model.

A. Development Layer

The initial part of the proposed MCM model is the development layer, which is further separated into two main sublayers: the application structure layer (microservices) and the code lifecycles layer. The microservice layer focuses on designing a good structure between each part of the application. The microservice layer seems to be as a pre-requisite for the code lifecycles. The development layer focuses on managing code changes and adopting CI principles to minimize code conflicts.

The code lifecycle layer consists of four steps. The first step verifies that the code is versioned correctly and that a stable version of the code has always been kept up to date in the main repository after the appropriate tagging. Developing a stable version of the code is the second step. The third step involves executing multiple test scenarios, starts after the build step is being successful. At this point, it is confirmed that the code satisfies the necessary functional and needed requirements. When the system's artifact is prepared for release, the last step in the build and validation software process is called "artifact preparation" This involves gathering the output required for the next release and deployment phases.

Routine 1 discusses the high-level routine for the code lifecycle sublayer.

Routine 1: Implement CI/CD Pipeline

Notations: SW // Software, VCS // Version control system, Repo // Repository, VAL // Validation, PR // Pull request, CONF // Confirmation, CI // Continuous integration, PL // Pipeline, TC // Test cases, Pkg // Package, Exec // Execute, CD // Continuous deployment
Input: SW code, Different types of TCs
Output: Deploy the succeeded container images to the target env
Steps:

```
1) Select VCS
2) Repo VAL
3) Code merge
If (PR= Yes) then
    While (review = accepted) do
        For (every merge CONF) do
            1) Push code
            2) Run CI PL
            3) Exec unit TCs
            4) Pkg the build artifact
            5) Exec extra TCs
            6) If (build pkg = accepted) then
                    Pkg the container image
                End
            7) If (CD env>=1) then
                    deploy pkg
                End
        End
    End
```

The procedures of this layer are concentrated on giving code and tests a transparent control structure. Selecting the most appropriate VCS is the initial stage, followed by defining the structure of the repository and its validation guidelines. Then get ready to create the different test cases required to get the CI/CD pipeline started. Finally, store the tested package and make it available for usage in deployment.

B. MLOps and Container Layer

The second layer of the MCM model is MLOps and Container layer. The MLOps and container layer is divided into two primary sublayers: the layer of MLOps comes first, and the layer of containers comes second.

MLOps layer includes the development of the MCM model, the pipelines construction, test preparation, configuring and adding some security gates throughout the MCM model.

Container layer, container as a service (CaaS), is the framework that relies on the idea of multiple containers. Therefore, each container in this layer will go through the same steps. The steps include container orchestration, cluster management, container security, image repository, routing services, and microservices mesh. The container's network configuration, security, and resource allocation are among the processes in this process, along with deployment, auto-scaling, health monitoring, migration, and load balancing.

The high-level routine for the MLOps layer is discussed in Routine 2. Routine 3 presents the container layer.

Routine 2: Implement MLOps Pipeline

Notations: DAQ // Data acquisition, DS // Data source, PL // Pipeline, DA // Data analysis, KPI // Key performance indicators, PERF // Performance, OP // Operational
Input: Model Data
Output: Deploy the succeeded ML model to the target env
Steps:

```
While (Business understanding = true) do
    For each (DAQ) do
        Define DS
        Define PL
        Define deployment env
        Exploration and cleaning
        End
    Model Security
    Model deployment
    For each (Model training) do
        Exploratory DA
        Define feature engineering
        Model training
        End
    For each (Model evaluation) do
        Model KPIs (PERF and OP)
        Model retrain
        End
    Model logging
    End
```

The MLOps layer consists of seven steps and the development of the MLOps pipeline is the main objective at this layer. The identification of the application type is the most crucial step because it serves as the base for the subsequent steps. The second step is understanding the application type which enables the identification and analysis of the received data. The data transformation step is the third step and begins after the data is received. The fourth step is setting up some security validation over the model pipeline and configuring all necessary deployment environments for the pipeline. The fifth step is the MCM's model training phase. The sixth step is the validation step. Depending on the outcomes of the MCM model validation step, the last step at MLOps layer which is the retraining step. The retraining step begins with some modifications to the MCM model and/or data to maximize the benefits. Furthermore, this step depends on the data in the MCM's model log file to keep track of the used data to be able to take the necessary action.

Routine 3: Implement Container Structure

Notations: STRTG // Strategy, CONFIG // Configuration, SP // Security policies, DR // Disaster recovery, HA // High availability, RS // Routing services
Input: Container STRTG
Output: CONFIG container architecture
Steps:

```
Container orchestration
CONFIG the container registry
    For each (Container) do
        Set container SP
        Define PL
        End
    Push container image
```

Along with the MLOps layer, the container layer which contains three steps starts. The first step is establishing the rules for container's interaction with the other containers. This process is known as container management and orchestration. The identification of the deployment strategy for the containers is the second step. The last step is defining all the image registries that depend on the various deployment environments.

C. Monitor Layer

The third layer of the MCM model is monitor layer. This layer includes MCM's model-based event logging, step-by-step monitoring, and the creation of dashboards for convenient monitoring to aid in decision-making on enhancements. The high-level routine for this layer is discussed in Routine 4 and it has been highlighted also in Routine 2 steps 6 and 7.

Routine 4: Monitoring

Notations: INIT: Initialize, PL // Pipeline, ML // Machine learning, REC // Record, CONFIG // Configuration
Input: Monitoring matrices
Output: Monitor PL, container, and ML model
Steps:

```
INIT monitoring
While (Logging = true) do
    For each (PL, container, and ML model) do
```

```
|  |  REC external logs
|  |  REC internal logs
|  | End
| End
CONFIG dashboards
```

In order to validate, enhance, and maintain any abnormal behavior over MCM's model layers, the monitoring layer contains various dashboards to represent multiple perspectives at various stages. The monitor layer saved external logs that record which data the MCM model is applied to, and internal logs that check the inner workings of the ML pipeline and debug problems. The monitoring layer makes it easier to understand and identify any failures, and act quickly to keep everything under control and in good condition. Moreover, the monitoring layer would guarantee a system with fewer development, deployment, and monitoring issues.

D. Tools and Automation Layer

The fourth layer of the MCM model is the Tools and Automation layer. This layer is necessary to support the MCM's model layers with one or more different tools to obtain reliable statistics and facilitate the transition from one step to another. Specifically, for provisioning and container cataloging, this layer uses a set of open-source and development automation tools to help in all MCM's model building and monitoring.

V. MCM MODEL DATASET

The dataset used in the of the proposed MCM model consists of 286 pipelines. The pipelines are divided to continuous integration, continuous testing, and continuous delivery pipelines that use 100 microservices. The pipelines serve over seven million users and run on Windows and Linux operating systems. The dataset contains 122 production environments, seven test and pre-production environments and 158 attributes spread across each layer of the suggested MCM model. Therefore, the size of the dataset is sufficient to validate the proposed MCM model and to investigate possible future enhancements. The dataset used by the MCM model collects and stores unstructured data from a variety of sources, including emails, business papers, source control software, and software engineers' feedback. A sample of the dataset is shown in Fig. 5 and a sample of the dataset and docker file are shown in Fig. 6.

commit ID	369ddcdb	cc3c53cf
number of reviewers on pull request	4	2
source control system type	Git	Git
application type	Web application	API
application language	.Net	Java
number of databases	2	1
number of APIs	0	3
agent status	Idel	Idel
agent version	3.240.1	3.240.1
pipeline triggers	Enable continuous integration	Batch changes while a build is in progress
branch filters	master	dev
pipeline created date	10/22/2018 12:35 AM	02/12/2023 5:13PM
pipeline changed date	03/18/2024 02:55 PM	06/15/2023 02:55 PM
path to publish	$(build.artifactstagingdirectory)	$(build.artifactstagingdirectory)
build status	Success	Canceled
Docker image to deploy	WebImage	JavaAPI
max artifact size	5 GB	5 GB
code analysis type	SonarQube	SonarQube
retention policy (d)	30	15
number of containers	1	1
operating system	Windows	Linux
container app environment	Workload profile	Workload profile
image tag	release	integration

Fig. 5. Example of the dataset representation.

```
# Set the locale
ENV LANG en_US.UTF-8
ENV LANGUAGE en_US:en
ENV LC_ALL en_US.UTF-8
RUN apt-get update && DEBIAN_FRONTEND=noninteractive apt-get install -y
locales

RUN locale-gen en_US.UTF-8 &&\
DEBIAN_FRONTEND=noninteractive dpkg-reconfigure locales

RUN DEBIAN_FRONTEND=noninteractive apt-get install -y fontconfig fontconfig-
config
#libvirt-clients not found?
RUN DEBIAN_FRONTEND=noninteractive apt-get install -y libvirt-bin virtinst
qemu-kvm libvirt-daemon libvirt-daemon-system qemu qemu-user wget shellcheck

RUN apt-get update && DEBIAN_FRONTEND=noninteractive apt-get -y upgrade

RUN modprobe kvm
RUN lsmod

EXPOSE 16509
#This didn't work?
RUN mknod /dev/kvm c 10 232

#Running in a container we can't access the socket?
RUN echo "listen_tls = 0" >> /etc/libvirt/libvirtd.conf; \
echo 'listen_tcp = 1' >> /etc/libvirt/libvirtd.conf; \
echo 'tls_port = "16514"' >> /etc/libvirt/libvirtd.conf; \
echo 'tcp_port = "16509"' >> /etc/libvirt/libvirtd.conf; \
echo 'auth_tcp = "none"' >> /etc/libvirt/libvirtd.conf

#RUN whoami
# root
#
#RUN pwd
# /
#
RUN ls -l /dev/
RUN ls -l /sys/module/
#RUN cat /sys/module/kvm_amd/parameters/nested
#RUN cat /sys/module/kvm_intel/parameters/nested
RUN cat /etc/issue

#setup ability to run kvm/qemu
VOLUME [ "/sys/fs/cgroup" ]
#RUN sed -i "/Service/a ExecStartPost=\/bin\/chmod 666 /dev/kvm"
/lib/systemd/system/libvirtd.service
RUN service libvirt-bin restart
RUN service qemu-kvm restart
RUN service libvirt-guests restart
RUN service virtlogd restart
RUN service --status-all

RUN mkdir tests

ADD docker-test.sh ./

#ADD /bin/chmod u+x docker-test.sh
#/bin/chmod no such file or directory
RUN ["chmod", "+x", "/docker-test.sh"]

#CMD ./docker-test.sh
```

Fig. 6. Example of the docker file.

VI. MCM MODEL CONFIGURATION

Every stage of the MCM's model preparation phase is covered in this section. This process consists of two primary preparation steps: building the application is the first main step, and setting up the ML model is the second.

A. Application Build

The migration phase and the automation phase are the two major stages of the application build process. The process depends on separating the application's functionality into discrete services according to their use cases or modules, and integration the DevOps with MLOps pipelines. Furthermore, the process focuses on using the containerization concept which makes it easier to manage and scale the system and makes it easier to automate the deployment of new features.

- Migration Phase from Monolithic Application to Containerized-Microservices Application

The goal of the migration process is to have a new, well-structured application to benefit from containers and DevOps tools in the next automation phase. Furthermore, the migration process identifies methods that will assist with securing the daily operation of the IT and reduce costs. Five primary stages made up the migration process, some of which had the subsequent sub-stages:

1) Evaluating the legacy application: The components of the MCM model were chosen with the need to address issues such as performance bottlenecks and reduce build and deployment durations.

2) Select a migration method: The Lift-and-Shift and Refactoring migration methods were chosen for the proposed MCM model.

3) Get the Legacy Application Ready for the Migration.

4) Data on migration: the GitOps tools, orchestration platform command-line interface (CLI), or orchestration platform web console can all be used to carry out deployment on the platform after migrating the application data. RedHat-OpenShift, and Kubernetes were chosen as the orchestration platforms for the proposed MCM model to scale resources for microservices and even to simply enable autoscaling. The platform CLI and Azure pipelines were used for the deployment execution.

5) Evaluate and deploy: once the application deployed on an orchestration platform, careful testing is also essential to confirm that the application performs as intended. The application can be launched to production if the testing is successful, which will complete the migration process.

- Automation Phase

This stage demonstrates how the refactored code "microservices" that were previously developed during the migration phase was built automatically based on Docker as shown in Fig. 7. One of the objectives of moving to microservices is to achieve the most benefit of the automated CI/CD pipeline, which enables a very seamless release process. In the case of an issue, the CI/CD pipeline system can initiate a fix or revert to an earlier version. Define checks at each stage using security scans, service level objectives (SLOs), service level indicators (SLIs), and service level agreements (SLAs).

Automate the changes between continuous integration, testing, delivery, and deployment.

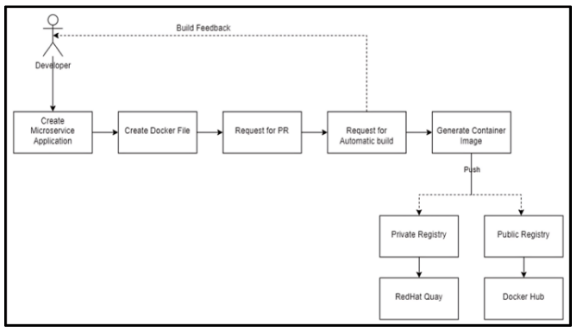


Fig. 7. Creation of container application and push the image to the registry.

The automation procedure was broken down into the following nine major phases:

1) Select a source control (SC) system. Git was the best SC system that met the requirements after several SC system types were compared.

2) Produce Docker files that specify each microservice's container contents.

3) Make a PR request.

4) Build and package every microservice, for Java projects, that have been developed and coded with Spring Boot and for .Net applications, which was created and developed with the .net framework versions 4.8.0, 4.8.1, and 6.

5) Build and produce Docker container images and verify if the image was locally created and accessible.

6) Launch the Docker containers and verify to see if the application within the container is operating properly.

7) Push the image to the registry, at this step, the locally produced image is pushed to the shared registry (RedHat Quay and/ or Docker Hub) using the image ID.

8) Initially, steps 3 through 8, are completed manually. Afterwards, CI/CD pipelines are set up using Azure DevOps. At this stage, each microservice has a CI pipeline to execute the code following an approved PR. The CT pipeline verifies the written code. The CD pipeline manages the deployment procedure for each environment, including integration, QC, security testing, UAT, load testing, packaging, and pre-production. A certain set of stakeholders must be informed and given permission before the created image is deployed on any deployment environment.

9) A new commit was added to the SC and step 3 was restarted to meet the additional requirements.

In conclusion, you can successfully containerize your application by following the previous steps, which include setting up the required tools, building the image, running the containers, and pushing the image to a registry. This will streamline the deployment and management process. Moreover, will encourage consistency and repeatability.

B. The Proposed MCM Model Build

The MCM model was built using a combination of recurrent neural network (RNN) architecture “Bidirectional LSTM (Bi- LSTM)” and a reinforcement learning algorithm “state–action–reward–state–action (SARSA)”. This combination integrates into application CI/CD pipelines and monitoring systems, enabling continuous optimization of container application performance. Further, by analyzing streaming data from container environments, the Bi-LSTM and SARSA can adapt to changing workload patterns and optimize resource usage dynamically. Combining the algorithms results in improved, robust, and shortened build and deployment times. The proposed MCM's model performance was evaluated using backpropagation through time (BPTT). The Adaptive Moment Estimation (Adam) optimizer was selected to improve the performance and accelerate the convergence of Bi-LSTM.

The proposed MCM’s model build process was divided into the seven main stages listed below:

1) Environment representation

a) Define the scope of the MCM’s model environment in terms of the performance and state of the containerized applications.

b) Normalize the metrics to make sure they were appropriate for input into MCM model and to guarantee consistent scaling.

2) Action scope

a) Specify a range of actions that the MCM model agent can accomplish to maximize container performance.

b) Discretization of the actions into a finite set.

3) Feedback function: provide the suggested MCM's agent a reward function that gives feedback based on the observed metrics, actions taken, and the obtained performance outcomes. The design of the reward function is broken down into three categories: positive, negative, and delayed feedback. Positive feedback rewards actions that improve container performance, like shorter build and deploy times, higher throughput, more efficient use of resources, and/or faster response times. Negative feedback penalizes actions that lead to inefficiency or performance degradation, like excessive resource consumption and downtime.

4) MCM's model agent: The proposed MCM's model agent depends on the SARSA's agent as shown in Fig. 8, to pick actions based on the current state and the observed feedback. MCM model learns an action-value function M (s, a) that estimates the expected cumulative feedback of taking an action (a) in the specified state (s) then a balanced exploration (random actions) and exploitation (best-known actions) strategy was combined with temporal difference learning in the SARSA algorithm to update the M-values.

5) Integrated MCM's model architecture: the MCM's model architecture is based on The Bi-LSTM design which consist of some LSTM layers, several hidden units per layer, and the size of input and output. System configurations, workload patterns, historical performance data, and other aspects were arranged into sequences and time-series representations to create the appropriately structured input data. The MCM model uses the input data to learn temporal patterns and anticipate future states. The SARSA agent then uses these predictions to make decisions on how best to optimize container performance and deployment durations. The schematic Bi-LSTM is shown in Fig. 9 and MCM's model integration architecture is shown in Fig. 10.

6) MCM model training: to ensure that gradients and errors are transmitted appropriately throughout time and to update the model weights while taking into consideration the sequential nature of the data, the MCM model was trained using BPTT. Further, Adam's optimizer was selected to modify the Bi-LSTM weights during training to minimize the loss function. Fig. 11 shows the flowchart for the training loop.

7) Evaluation and deployment

a) A validation dataset and container assessment measures, such as response times, throughput, resource utilization, and build/deploy durations, were utilized to assess the MCM model algorithm's performance.

b) Embed the trained MCM's model algorithm in the CI/CD pipelines and containerized application deployment pipelines to automate the entire deployment process and guarantee a smooth integration with the current DevOps workflows.

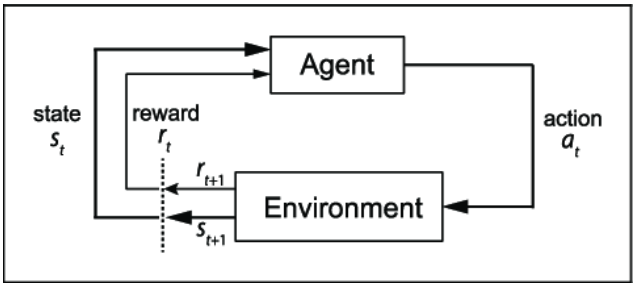


Fig. 8. MCM's model agent environment interaction.

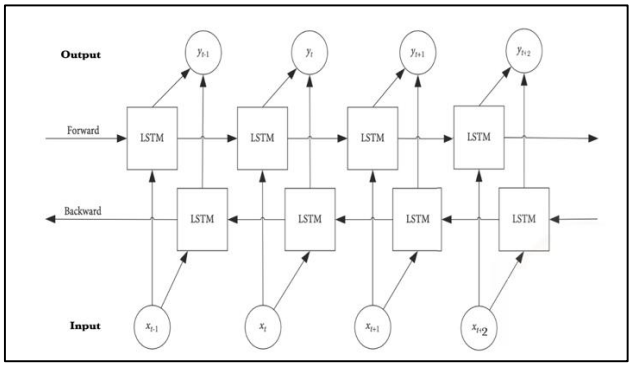


Fig. 9. Schematic Bi-LSTM's architecture.

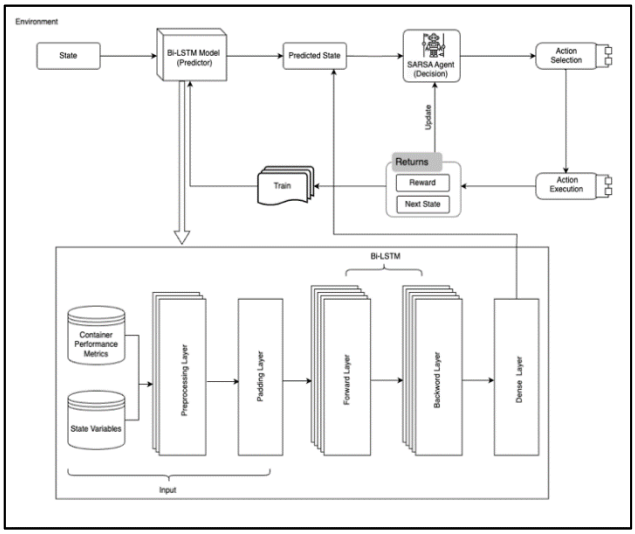


Fig. 10. Integrated MCM's model architecture

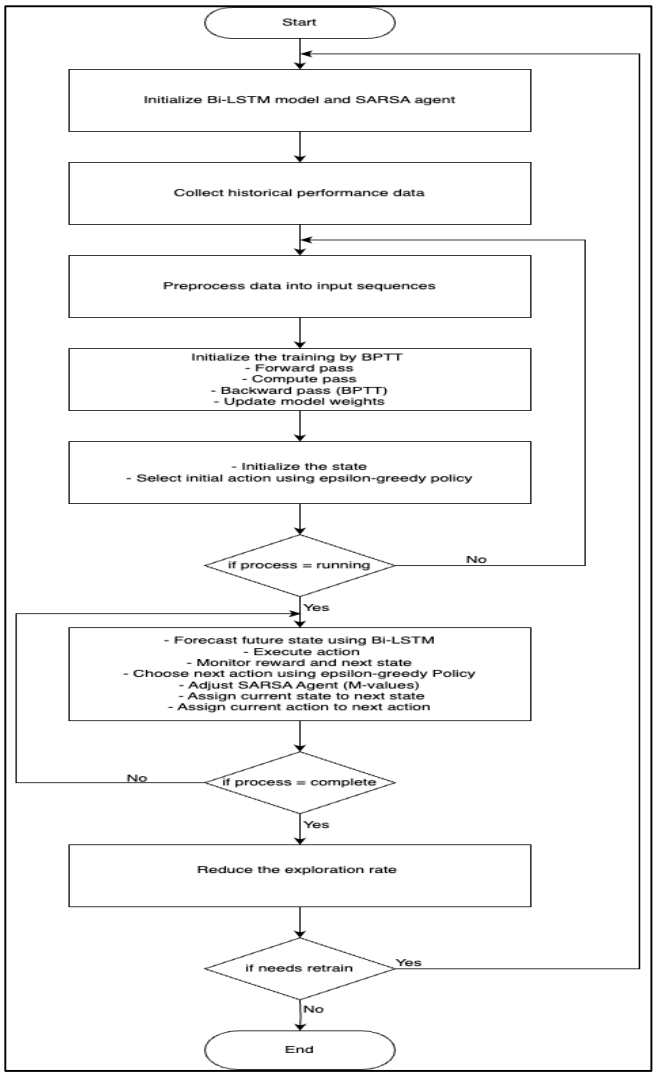


Fig. 11. MCM's model training flowchart.

VII. MCM Model Results

The outcomes of a successful progress visualization during MCM model training to improve the deployment time are presented in Fig. 12. The improvements in build and deployment durations are shown in Fig. 13.

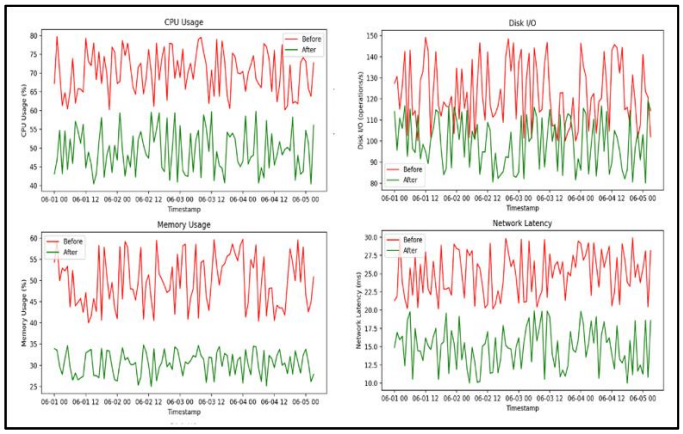


Fig. 12. MCM's model resource usage optimization.

Fig. 12 demonstrates optimization of the resource usage of the containers. It can be noticed that the containers enhancements made by the MCM model compared to the previously employed methodologies for CUP usage is up to 38.25%, for disk I/O is up to 39.20%, for memory usage is up to 50.77% and for network latency is up to 58.37%.

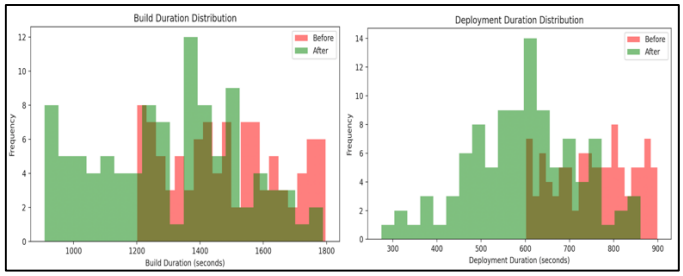


Fig. 13. MCM's model build and deployment duration.

Fig. 13 observes the improvements in seconds of the build and deployment frequencies across the different environments, which are for a build duration up to 13% and for a deployment duration up to 24.55%.

VIII. MCM Model Discussion

The literature research revealed several barriers to the use of MLOps. These unresolved concerns fall into three categories: organizational, ML system, and operational.

A. Organizational Challenges

A common issue in organizational settings is the mentality and culture of data science practice [31]. The study's findings indicate that to effectively develop, implement, and monitor machine learning systems, there require a culture shift away from model-driven machine learning and toward a discipline that is system-oriented. This can be accomplished by placing more emphasis on the data-related activities that take place before the creation of the ML model. Furthermore, when designing ML products, roles involved in these activities should have a system-focused view and must therefore be a group process as this is challenging since teams usually operate in silos rather than collaborative environments also the specialized terminologies and varying degrees of knowledge further complicate communication. Moreover, MLOps demands a wide range of skills and specialized roles. Because there aren't enough highly qualified professionals to fill these positions, particularly in the fields of architects, data engineers, ML engineers, and DevOps engineers [32] [33] [34]. As MLOps is often not included in data science courses, this is relevant to the training that future workers will need [35].

B. ML System Challenges

It might be difficult to develop MLOps systems to accommodate changing demands, especially when it comes to the ML training and monitoring procedures [35]. This is a result of potentially massive and unpredictable data [36], which makes it challenging to correctly forecast the necessary infrastructure resources (CPU, RAM, and GPU), and calls a high level of flexibility when it comes to the scalability of the containers [35] [37].

C. Operational Challenges

ML is challenging to execute manually because of the numerous software and hardware stacks and their interrelationships. Thus, reliable automation is required to produce numerous artifacts, which require solid governance [35] [38] [39] [32] [40] [41]. Versioning of the data, model, and code is also necessary to guarantee reliability and reproducibility [32] [2]. Finally, because there are so many parties and components involved, it can be difficult to handle a possible support request (for example, by identifying the root cause). Moreover, failures might result from a combination of ML infrastructure and software [37] [42], making it essential to monitor each phase and collect as much data as possible to aid in making timely decisions.

Therefore, this research proposed a model (MCM) which depends on the multi-container architecture and microservices principles that applied to solve the mentioned barriers by building and deploying the stages of the application development lifecycle. The MCM reduced the requirement for re-developing and re-deploying software applications while also improving the performance of software releases. The MCM model enhances the build duration cycles and software deployment cycles ratio by employing MLOps.

IX. Conclusions and Future Work

More machine learning systems than ever before are being developed as a result of the growing demand to innovate. Higher monitoring and analysis skills are required for ML models. However, only a few of these proofs of concept move forward to deployment to production. Furthermore, in the real world, data scientists are still managing ML operations largely manually. These issues are addressed by the Machine Learning Operations (MLOps) paradigm. Moreover, according to the linked publications, there are no studies that concentrate on monitoring MLOps applications, especially those that rely on multi-container and microservices design. Therefore, this research proposed a model (MCM) which depends on the multi-container architecture and microservices principles that applied to the build and deploy stages of the application development lifecycle. The developed MCM model used to increase the number of software deployments across a variety of environments. Further the proposed MCM improved the software release performance and decreased the need for re-developing and re-deploying software applications. By utilizing MLOps, the suggested MCM model improves the software deployment cycles ratio by up to 24.55% and build duration cycles by up to 13%. This was useful in directing different IT teams towards the areas of monitoring ML model's features by using MLOps. Furthermore, the research recommended four routines for each layer of the suggested MCM model, described how each layer will be developed. As future work, more experimental work is also needed to assess the MLOps pipelines and see how they might affect the overall software development cycle. The MCM model needs to be implemented on different data sets and monitor its efficiency. More experiments to compare the performance of the MCM model algorithm against baseline approaches or alternative optimization strategies are needed.

References

[1] F. Calefato, F. Lanubile, and L. Quaranta, A Preliminary Investigation of MLOps Practices in GitHub, vol. 1, no. 1. Association for Computing Machinery, 2022. doi: 10.1145/3544902.3546636.

[2] D. Kreuzberger, N. Kühl, and S. Hirschl, "Machine Learning Operations (MLOps): Overview, Definition, and Architecture," 2022, [Online]. Available: http://arxiv.org/abs/2205.02302.

[3] S. Makinen, H. Skogstrom, E. Laaksonen, and T. Mikkonen, "Who needs MLOps: What data scientists seek to accomplish and how can MLOps help?," Proc. - 2021 IEEE/ACM 1st Work. AI Eng. - Softw. Eng. AI, WAIN 2021, pp. 109–112, 2021, doi: 10.1109/WAIN52551.2021.00024.

[4] I. Karamitsos, S. Albarhami, and C. Apostolopoulos, "Applying devops practices of continuous automation for machine learning," Inf., vol. 11, no. 7, pp. 1–15, 2020, doi: 10.3390/info11070363.

[5] Y. Liu, Z. Ling, B. Huo, B. Wang, T. Chen, and E. Mouine, "Building A Platform for Machine Learning Operations from Open Source Frameworks," IFAC-PapersOnLine, vol. 53, no. 5, pp. 704–709, 2020, doi: 10.1016/j.ifacol.2021.04.161.

[6] L. Baier and S. Seebacher, "Challenges in the Deployment and," 27th Eur. Conf. Inf. Syst., no. May, pp. 1–15, 2019, [Online]. Available: https://aisel.aisnet.org/ecis2019_rp/163/

[7] D. A. Tamburri, "Sustainable MLOps: Trends and Challenges," Proc. - 2020 22nd Int. Symp. Symb. Numer. Algorithms Sci. Comput. SYNASC 2020, pp. 17–23, 2020, doi: 10.1109/SYNASC51798.2020.00015.

[8] O. Spjuth, J. Frid, and A. Hellander, "The machine learning life cycle and the cloud: implications for drug discovery," Expert Opin. Drug Discov., vol. 16, no. 9, pp. 1071–1079, 2021, doi: 10.1080/17460441.2021.1932812.

[9] E. Calikus, Self-Monitoring using Joint Human- Machine Learning : Algorithms and Applications, no. 69.

[10] T. Schröder and M. Schulz, "Monitoring machine learning models: a categorization of challenges and methods," Data Sci. Manag., vol. 5, no. 3, pp. 105–116, 2022, doi: 10.1016/j.dsm.2022.07.004.

[11] L. Cardoso Silva et al., "Benchmarking Machine Learning Solutions in Production," Proc. - 19th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2020, no. March, pp. 626–633, 2020, doi: 10.1109/ICMLA51294.2020.00104.

[12] P. Liang et al., "Automating the training and deployment of models in MLOps by integrating systems with machine learning", Proceedings of the 2nd International Conference on Software Engineering and Machine Learning, 2024, doi: 10.54254/2755-2721/67/20240690.

[13] C. Wu, E. Haihong, and M. Song, "An Automatic Artificial Intelligence Training Platform Based on Kubernetes," ACM Int. Conf. Proceeding Ser., pp. 58–62, 2020, doi: 10.1145/3378904.3378921.

[14] B. Karlaš et al., "Building Continuous Integration Services for Machine Learning," Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., no. November, pp. 2407–2415, 2020, doi: 10.1145/3394486.3403290.

[15] P. Ruf, M. Madan, C. Reich, and D. Ould-Abdeslam, "Demystifying mlops and presenting a recipe for the selection of open-source tools," Appl. Sci., vol. 11, no. 19, 2021, doi: 10.3390/app11198861.

[16] Y. Liu, "Understanding MLOps : a Review of " Practical Deep Learning at Scale with Understanding MLOps : a Review of ' Practical Deep Learning at Scale with MLFlow ' by Yong Liu," no. July, 2022, doi: 10.13140/RG.2.2.21031.83369.

[17] Z. Shoieb, L. Abdelhamid, M. Abdelfattah, "Enhancing Software Deployment Release Time Using DevOps Pipelines", IJSER, vol.11, no. 3, 2020, ISSN: 2229-5518.

[18] M. Rowse and J. Cohen, "A survey of DevOps in the South African software context," Proc. Annu. Hawaii Int. Conf. Syst. Sci., vol. 2020-Janua, pp. 6785–6794, 2021, doi: 10.24251/hicss.2021.814.

[19] A. Sajid et al., "AI-Driven Continuous Integration and Continuous Deployment in Software Engineering" 2nd International Conference on Disruptive Technologies (ICDT), 2024.

[20] R. Subramanya, S. Sierla, and V. Vyatkin, "From DevOps to MLOps: Overview and Application to Electricity Market Forecasting," Appl. Sci., vol. 12, no. 19, 2022, doi: 10.3390/app12199851.

[21] B. Mayumi, A. Matsui, and D. H. Goya, "Applying DevOps to Machine Learning Processes : A Systematic Mapping," 2019.

[22] T. Zheng et al., "NCCMF: Non-Collaborative Continuous Monitoring Framework for Container-Based Cloud Runtime Status", Computers, Materials & Continua, 2024, doi: 10.32604/cmc.2024.056141

[23] P. Agrawal and N. Rawat, "Devops, A New Approach to Cloud Development Testing," IEEE Int. Conf. Issues Challenges Intell. Comput. Tech. ICICT 2019, 2019, doi: 10.1109/ICICT46931.2019.8977662.

[24] N. Hewage and D. Meedeniya, "Machine Learning Operations: A Survey on MLOps Tool Support," no. February, 2022, doi: 10.48550/arXiv.2202.10169.

[25] S. Alla and S. K. Adari, Beginning MLOps with MLFlow. 2021. doi: 10.1007/978-1-4842-6549-9.

[26] G. Recupito et al., "A Multivocal Literature Review of MLOps Tools and Features," no. July, pp. 84–91, 2023, doi: 10.1109/seaa56994.2022.00021.

[27] L. E. L. B, I. Crnkovic, R. Ellinor, and J. Bosch, "From a Data Science Driven Process to a Continuous Delivery Process for Machine Learning Systems," Proceedings- PROFES- 21st Int. Conf., vol. 1, 2020.

[28] G. E. De Velp, E. Rivière, and R. Sadre, "Understanding the performance of container execution environments," WOC 2020 - Proc. 2020 6th Int. Work. Contain. Technol. Contain. Clouds, Part Middlew. 2020, no. 37, pp. 37–42, 2020, doi: 10.1145/3429885.3429967.

[29] C. Segarra et al., "Serverless Confidential Containers: Challenges and Opportunities" 2024.

[30] B. Burns, "Design patterns for container-based distributed systems".

[31] Z. Shen et al., "X-Containers: Breaking Down Barriers to Improve Performance and Isolation of Cloud-Native Containers," Int. Conf. Archit. Support Program. Lang. Oper. Syst. - ASPLOS, pp. 121–135, 2019, doi: 10.1145/3297858.3304016.

[32] E. Summary, "PRINCIPLES OF CONTAINER-BASED".

[33] R. Madhumathi, "The Relevance of Container Monitoring Towards Container Intelligence," 2018 9th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2018, pp. 1–5, 2018, doi: 10.1109/ICCCNT.2018.8493766.

[34] P. Liu and J. Guitart, "Performance comparison of multi-container deployment schemes for HPC workloads: an empirical study," J. Supercomput., vol. 77, no. 6, pp. 6273–6312, 2021, doi: 10.1007/s11227-020-03518-1.

[35] J. Brier and lia dwi jayanti, "DevSecOps of Containerization," vol. 21, no. 1, pp. 1–9, 2020, [Online]. Available: http://journal.um-surabaya.ac.id/index.php/JKM/article/view/2203

[36] H. Gantikow, C. Reich, M. Knahl, and N. Clarke, "Rule-Based Security Monitoring of Containerized Environments," Commun. Comput. Inf. Sci., vol. 1218 CCIS, pp. 66–86, 2020, doi: 10.1007/978-3-030-49432-2_4.

[37] A. Mahesar et al., "Efficient microservices offloading for cost optimization in diverse MEC cloud networks". *J Big Data, vol*. 11, no. 123, 2024. https://doi.org/10.1186/s40537-024-00975-w

[38] Z. Zhong, M. Xu, M. A. Rodriguez, C. Xu, and R. Buyya, "Machine Learning-based Orchestration of Containers: A Taxonomy and Future Directions," ACM Comput. Surv., vol. 54, no. 10s, pp. 1–35, 2022, doi: 10.1145/3510415.

[39] R. Miñón, J. Diaz-De-arcaya, A. I. Torre-Bastida, and P. Hartlieb, "Pangea: An MLOps Tool for Automatically Generating Infrastructure and Deploying Analytic Pipelines in Edge, Fog and Cloud Layers," Sensors, vol. 22, no. 12, 2022, doi: 10.3390/s22124425.

[40] M. Testi et al., "MLOps: A Taxonomy and a Methodology," IEEE Access, vol. 10, no. June, pp. 63606–63618, 2022, doi: 10.1109/ACCESS.2022.3181730.

[41] S. Moreschini, F. Lomio, D. Hastbacka, and D. Taibi, "MLOps for evolvable AI intensive software systems," Proc. - 2022 IEEE Int. Conf. Softw. Anal. Evol. Reengineering, SANER 2022, no. January, pp. 1293–1294, 2022, doi: 10.1109/SANER53432.2022.00155.

[42] G. Bou Ghantous and A. Q. Gill, Evaluating the DevOps Reference Architecture for Multi-cloud IoT-Applications, vol. 2, no. 2. Springer Singapore, 2021. doi: 10.1007/s42979-021-00519-6.

Deep Learning Approach in Complex Sentiment Analysis: A Case Study on Social Problems

Bambang Nurdewanto[1], Kukuh Yudhistiro[*2,] , Dani Yuniawan[3], Himawan Pramaditya[4], Mochammad Daffa Putra Karyudi[5], Yulia Natasya Farah Diba Arifin[6] , Puput Dani Prasetyo Adi[7]

Faculty of Information Technology, Universitas Merdeka Malang, Malang City, East Java, Indonesia[1, 2, 3, 4, 5, 6]
National Research and Innovation Agency (BRIN), Bandung Indonesia[7]

***Abstract*—This scholarly investigation examines the utilization of artificial intelligence (AI) technology in the analysis and resolution of intricate societal challenges in many countries. The originality of this study resides in the employment of deep learning algorithms, particularly Convolutional Neural Network (CNN), to execute sentiment analysis with an elevated degree of complexity. The examination encompasses three principal dimensions of sentiment: Sentiment, Tone, and Object, with the intention of offering profound insights into public perceptions regarding various social challenges. The fundamental sentiment is categorized into three classifications: Positive, Neutral, and Negative. Moreover, the Tone analysis introduces an additional layer of comprehension that encompasses Support, Suggestion, Criticism, Complaint, and Others, thereby delineating a more precise communicative context. The Object dimension is employed to ascertain the target of the sentiment, whether it pertains to an Individual, Organization, Policy, or other entity. This inquiry applied the analysis to several clusters of social issues, including Poverty and Economic Disparity, Health and Wellbeing, Education and Literacy, Violence and Security, as well as Environment and Social Life. The findings are anticipated to aid the government in devising policies that are more effective and responsive to the exigencies of society, through an enhanced understanding of public sentiment.**

***Keywords*—Component; sentiment analysis; deep learning; artificial intelligence; social case**

I. INTRODUCTION

Sentiment analysis plays a critical role in understanding and quantifying the emotions or sentiments expressed in textual data, such as reviews, comments, and social media posts. It is an essential tool for businesses and organizations aiming to comprehend public opinion, improve customer service, and make data-driven decisions in product development and marketing strategies [10]. However, traditional sentiment analysis methods struggle with capturing the complexities of language, including sarcasm, ambiguous expressions, and nuanced opinions [1]. This limitation has led to a growing need for more sophisticated techniques to accurately analyze sentiment, especially when dealing with large volumes of unstructured data from sources like social media [7].

To address these challenges, recent research has increasingly employed deep learning models, which offer significant improvements over traditional methods in sentiment classification. Neural networks, particularly deep learning architectures such as capsule-based RNNs and Graph Convolutional Networks (GCNs), have shown remarkable success in handling complex textual data [6] [14]. These models not only enhance sentiment classification but also allow for the integration of multimodal data—such as text, audio, and visual cues—to capture deeper insights, especially in the case of complex emotions like sarcasm [8] [12]. Moreover, transfer learning techniques have been implemented to overcome data scarcity challenges, enabling models to be fine-tuned for smaller, domain-specific datasets [11] [16].

This study contributes to the field by employing Convolutional Neural Networks (CNNs), a deep learning methodology traditionally associated with image processing, for sentiment analysis [9]. CNNs have demonstrated effectiveness in analyzing text by treating it as spatial data, extracting local features, and offering a more straightforward yet efficient approach compared to other models [18]. By leveraging CNNs, this research aims to improve sentiment analysis performance, particularly in understanding nuanced social issues [20], and to provide more precise and actionable insights [21] [22].

In line with the abstract, this research focuses on three principal dimensions of sentiment analysis: Sentiment, Tone, and Object. Sentiment is categorized into three primary classifications—Positive, Neutral, and Negative—while Tone introduces additional layers of interpretation, including Support, Suggestion, Criticism, Complaint, and Other. The Object dimension further refines the analysis by identifying the target of the sentiment, such as an Individual, Organization, or Policy. This comprehensive approach enables a more detailed understanding of public perceptions regarding various social challenges.

Specifically, the research addresses key societal clusters such as Poverty and Economic Disparity, Health and Wellbeing, Education and Literacy, Violence and Security, and Environment and Social Life. The insights generated through this analysis are expected to support government agencies in developing policies that are more responsive to public sentiment and aligned with societal needs [23]. Through the integration of CNNs in sentiment analysis, this study offers a novel approach to addressing the limitations of traditional methods and aims to enhance decision-making processes by providing deeper insights into public opinion on critical social issues.

II. LITERATURE REVIEW

A. Sentiment Analysis

Sentiment analysis plays a crucial role in shaping social justice movements by providing insights into public opinion and

emotional responses [24]. By analyzing sentiments expressed on social media platforms, activists can better understand the dynamics of support and opposition [30], which can inform strategies for advocacy and mobilization [27].

Impact on Public Sentiment: Black Lives Matter Movement: A study utilizing a BERT model analyzed over one million tweets related to the BLM movement, revealing themes of social justice and police brutality [29]. Positive sentiments were linked to significant events, guiding activists in their messaging and outreach efforts [13]. Sentiment assessment, akin to employing BERT on Twitter datasets pertaining to Black Lives Matter, elucidates public sentiment regarding social justice initiatives, accentuating motifs and occurrences linked with affirmative or adverse sentiment [25]. The BERT model, meticulously adjusted for Twitter sentiment, surpassed alternative models in performance. Retweet and lexical count frequencies underscored significant BLM motifs/occurrences.

EndSARS Movement: The sentiment analysis of the #EndSARS protests in Nigeria highlighted eight basic emotions, providing insights into public sentiment that could inform governmental responses and future activism strategies [14]. Sentiment analysis, as demonstrated in the research concerning the EndSARS demonstration in Nigeria, facilitates the comprehension of collective sentiments, directing governmental reactions, and potentially mitigating social turbulence in forthcoming movements [26]. Recognized eight fundamental emotions articulated during the EndSARS demonstration. Suggested methodologies for the government to tackle subsequent activist movements.

Emotional Dynamics: Chinese Online Movements: Research indicates that emotions like anger and anxiety are pivotal in mobilizing online social movements [28]. Understanding these sentiments can help predict movement trajectories and prevent potential social unrest [15]. Sentiment evaluation in digital social movements, particularly concentrating on outrage and apprehension, facilitates comprehension of sentiment progression to avert societal calamities, augmenting the efficacy of social justice campaigns. The suggested BERT-based framework surpasses foundational models. A sentiment evaluation dataset for Chinese digital social movements has been established.

Broader Implications: Social Justice and Well-being: The correlation between feelings of social justice and subjective well-being suggests that sentiment analysis can also reflect broader societal perceptions, influencing public policy and community engagement [19]. The investigation examines the robust association between perceptions of social equity and subjective welfare, proposing that sentiments of justice within a community affect personal life contentment and well-being [31]. Welfare is impacted by financial resources, social connections, and perceptions of equity. Significant association between life contentment and societal justice interpretations.

While sentiment analysis offers valuable insights, it is essential to consider the potential for misinterpretation or manipulation of data, which can skew public perception and impact the effectiveness of social justice movements.

Previous CNN research related to sentiment analysis for social problems, Sentiment analysis using deep learning, particularly Convolutional Neural Networks (CNNs), has gained traction in addressing social issues through the analysis of public sentiment on platforms like Twitter and Reddit. This approach leverages advanced methodologies to extract meaningful insights from vast datasets, enhancing understanding of societal concerns [32].

Deep Learning Techniques in Sentiment Analysis: CNNs and RNNs: Recent studies highlight the effectiveness of CNNs and hybrid models like CRDC (Capsule with Deep CNN and Bi-structured RNN), achieving accuracy rates up to 98.02% in sentiment classification tasks [6] The manuscript presents an innovative hybrid methodology, CRDC, amalgamating Capsule, Deep CNN, and Bi-RNN for sentiment analysis, surpassing conventional CNN or RNN frameworks with elevated precision across diverse datasets. Although the Capsule-based RNN methodology exceeds the performance of CNN or RNN models, the CRDC model demonstrates exceptional efficacy in sentiment analysis.

LSTM Models: While CNNs are prominent, LSTM models have also shown significant results, with an 88.7% accuracy in analyzing sentiments regarding animal testing (Ismail, 2024). This document explores the realm of affective assessment utilizing Long Short-Term Memory (LSTM) networks rather than relying on Convolutional Neural Networks (CNN). It does not tackle the application of CNN in social contexts. A preponderance of tweets articulated adverse sentiment regarding animal experimentation. The LSTM architecture attained an accuracy of 88.7% in sentiment categorization.

Applications in Social Contexts: Crisis Response and Governance: AI/ML tools, including sentiment analysis, are utilized to gauge public opinion on governance issues, as demonstrated in a study analyzing 114,390 Reddit comments, achieving an accuracy of 89% with the SiEBERT model [8]. The manuscript centers on affective evaluation employing SiEBERT, a pre-trained transformer architecture, attaining 89% precision, signifying the efficacy of deep learning artificial intelligence in comprehending the apprehensions of citizens via social media sentiment assessment. BERTopic with K-means attained the utmost coherence rating.

Consumer Insights: Automated sentiment analysis models, such as ASASM-HHODL, have been developed to process social media data effectively, enhancing brand monitoring and customer feedback analysis [5]. The manuscript advocates for an Automated Sentiment Evaluation in Social Media (ASASM) employing Harris Hawks Optimization and Deep Learning (HHODL) methodologies, concentrating on Attention-oriented BiLSTM for sentiment categorization, without explicitly referencing CNN. The ASASM-HHODL framework surpasses contemporary sentiment analysis methodologies. Deep Learning models autonomously derive characteristics from datasets for sentiment evaluation.

While deep learning models, particularly CNNs, show promise in sentiment analysis, challenges remain in data preprocessing and model optimization, necessitating ongoing research to refine these techniques for broader applications.

How do Convolutional Neural Networks impact sentiment analysis?, Convolutional Neural Networks (CNNs) significantly enhance sentiment analysis by effectively capturing local patterns in text data, improving classification accuracy and interpretability. Their ability to process n-grams and semantic features allows for nuanced sentiment detection, as evidenced by various studies [33].

CNNs facilitate understanding of sentiment classification through the analysis of convolutional filters, revealing that certain parts of speech (POS) are more influential in determining sentiment polarity. The preference for shorter n-grams in negative sentiment classification highlights the model's focus on critical linguistic features [4]. CNNs impact sentiment analysis by revealing the relevance of certain parts of speech tags and the preference for shorter n-grams in classifying negative sentiment sequences, indicating the potential for smaller architectures.

Hybrid Models and Performance: Integrating CNNs with other architectures, such as BERT and BiLSTM, enhances the extraction of both local and global semantic features, achieving high accuracy rates (e.g., 92.35%) in sentiment classification tasks [2]. The analysis showcases a cutting-edge hybrid system that integrates two pathways in neural networks, employing an attention mechanism to expertly combine Convolutional Neural Networks (CNNs) and Bidirectional Long Short-Term Memory networks (BiLSTMs), hence significantly boosting sentiment analysis via the proficient acquisition of both local and global semantic elements. The suggested framework attains a sentiment classification accuracy of 92.35%. Moreover, the outlined system reaches an F1 score of 91.59% regarding sentiment classification.

Graph-based approaches, like hierarchical dual graph convolutional networks, further improve aspect-based sentiment analysis by effectively modeling syntactic and semantic dependencies [17]. Hierarchical dual graph convolutional network (HD-GCN) enhances sentiment analysis by extracting syntactic and semantic information, outperforming state-of-the-art models on ABSA tasks, as shown in the study.

While CNNs excel in sentiment analysis, their black-box nature raises concerns about interpretability and understanding of internal representations, which can hinder trust in automated systems [15]. Convolutional Neural Networks (CNNs) play a crucial role in sentiment analysis by extracting features from text data, enhancing sentiment classification accuracy, and addressing interpretability challenges in neural network models.

III. Research Method

A. Sentiment Aspects and the Labels

Sentiment aspect prediction refers to the process of identifying and categorizing various components of sentiment expressed within a text. It involves breaking down the text to determine not just the overall sentiment (positive, neutral, negative) but also other relevant aspects that give more context to the sentiment. Here's a breakdown of how predictions for each sentiment aspect typically work: Basic Sentiment: [Positive, Neutral, Negative], According to Pang and Lee (2008), basic sentiment analysis focuses on classifying text into these three main categories. This classification is a starting point in many sentiment analysis applications.

- *Positive*: Indicates a supportive or optimistic attitude toward the topic.
- *Neutral*: Indicates a neutral or impartial attitude.
- *Negative*: Indicates an opposing or pessimistic attitude toward the topic.

Furthermore, Tone: [Support, Suggestion, Criticism, Complaint, Others], According to Liu (2012), understanding tone in sentiment analysis provides deeper insights into how an entity or issue is perceived, not only from the polarity perspective but also from the communication context.

- *Support*: Text that shows approval or support for an issue or entity.
- *Suggestion*: Text that provides recommendations or solutions to a specific problem.
- *Criticism*: Text that expresses negative opinions or disagreement.
- *Complaint*: Text that conveys dissatisfaction or problems encountered.

Furthermore, Object: [Individual, Organization, Policy, Others]. According to Balahur et al. (2010), identifying the object in sentiment analysis helps understand the main target of the expressed sentiment, which is crucial in the context of social issues.

- *Individual*: Text referring to a person or a group of individuals.
- *Organization*: Text referring to organizational entities such as companies, institutions, or governments.
- *Policy*: Text referring to specific regulations, laws, or policies.

Furthermore, Cluster: [Poverty and Economic Inequality, Health and Well-being, Education and Literacy, Violence and Security, Environment and Social Life, Others], According to Cambria et al. (2013), clustering text based on topics or issues provides clearer context about the area of concern being analyzed, enabling deeper understanding and more targeted actions.

- *Poverty and Economic Inequality*: Text related to economic issues and social disparities.
- *Health and Well-being*: Text related to health issues, healthcare services, and social well-being.
- *Education and Literacy*: Text related to education systems, access to education, and literacy.
- *Violence and Security*: Text related to issues of violence, crime, and public security.
- *Environment and Social Life*: Text related to environmental issues and social interactions.

B. Data Collection

One essential parameter is Data Collection (see Fig. 1 and Fig. 2), Dataset that has been labeled for basic sentiment aspects

(Positive, Neutral, Negative). The dataset has been labeled for the Tone sentiment aspects (Support, Suggestions, Criticism, Complaints, Others).

```
import pandas as pd

data = {
 'text': [
  'I strongly support this program.',
  'This program should be improved.',
  'I am not satisfied with this service',
  'This service is very disappointing',
  'There is nothing special about this program'
 ],
 'label': ['Support', 'Suggestions', 'Criticism', 'Complaints', 'Other']
}

df = pd.DataFrame(data)
```

Fig. 1. Data Collection code 1.

```
import pandas as pd

data = {
 'text': [
  'I strongly support this program.',
  'This program should be improved.',
  'I am not satisfied with this service',
  'This service is very disappointing',
  'There is nothing special about this program'
 ],
 'label': ['Support', 'Suggestions', 'Criticism', 'Complaints', 'Other']
}

df = pd.DataFrame(data)
```

Fig. 2. Data Collection code 2.

Furthermore, The same method for the Tone aspect and basic sentiment is also applied to the Object sentiment aspects (Individual, Organization, Policy) and Cluster aspect (Poverty and Economic Inequality, Health and Well-being, Education and Literacy, Violence and Security, Environment and Social Life, Others).

C. Text Preprocessing

In this Text Preprocessing process there are several things that are done, namely Cleaning and preparing text data to be ready for analysis. And Steps. Steps consist of several parts, namely:

- Tokenization : Breaking down the text into smaller units, such as words or phrases. Example: "Poverty in this area is worsening due to economic inequality" -> ["Poverty", "in", "this", "area", "is", "worsening", "due", "to", "economic", "inequality"]
- Punctuation Removal: Removing unnecessary punctuation marks. Example: ["Poverty", "in", "this", "area", "is", "worsening", "due", "to", "economic", "inequality"]
- Stop Words Removal: Removing common words that do not carry significant meaning, such as "and," "or," "is." Example: ["Poverty", "area", "worsening", "due", "economic", "inequality"]
- Stemming or Lemmatization : Converting words to their base form.Example: ["Poverty", "area", "worsen", "due", "economic","inequality"]

D. Feature Extraction

Furthermore, Feature Extraction: Identifying and extracting important features from the text that will be used for classification. Steps: Using Term Frequency-Inverse Document Frequency (TF-IDF) to Calculate the importance of a word in a document compared to the entire corpus. Example: ["Poverty", "area", "worsening", "inequality", "economic"] -> [0.2, 0, 0.4, ..., 0.3, 0.5]

E. Model Development Using CNN

Training a machine learning or deep learning model to recognize cluster patterns in the text. Steps:

- Training the model using a Convolutional Neural Networks (CNN) architecture to recognize cluster patterns in the text.
- Training the model using data labeled.
- Model Validation: Evaluating the model using validation data to ensure its performance.

F. Prediction for Each Sentiment Aspect

Load the trained model: This step involves loading a pre-trained machine learning or deep learning model that has been specifically trained to analyze sentiment aspects. The model is typically saved in a file after training and can be loaded using various libraries such as TensorFlow or PyTorch. Loading the trained model allows us to utilize its learned parameters and weights for predicting sentiment aspects in new text data.

Make cluster predictions: Once the trained model is loaded, it can be used to make predictions on new, unseen data. This step involves feeding the input text data into the model to predict the specific clusters or categories of sentiment aspects. The output will indicate which sentiment cluster the text belongs to, such as "Poverty and Economic Inequality," "Health and Well-being," or other predefined clusters.

G. Evaluation and Validation

Evaluating the performance of the model using metrics such as accuracy, precision, recall, and F1-score. Steps:

- Confusion Matrix : Analyzing the model's prediction errors.
- Cross-validation: Performing cross-validation to ensure the model is not overfitting.

IV. Result and Discussion

A. Text Preprocessing

The implementation of the algorithm in the methodology section is implemented using the Python programming language with the necessary libraries such as TensorFlow.Text Preprocessing: For as example, in the following text as data input: 'This program aims to reduce poverty and improve community welfare.' Steps: Tokenization (see Fig. 3), Breaking down the text into smaller units, such as words or phrases. Fig. 4 shows text processing code.

```
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
from sklearn.preprocessing import LabelEncoder

# Tokenisasi teks
tokenizer = Tokenizer(num_words=5000)
tokenizer.fit_on_texts(df['text'])
sequences = tokenizer.texts_to_sequences(df['text'])

# Padding sequences
max_length = 50
padded_sequences = pad_sequences(sequences, maxlen=max_length)

# Encode label
label_encoder = LabelEncoder()
labels = label_encoder.fit_transform(df['label'])
```

Fig. 3. Tokenization code.

output: {'program': 1, 'aim': 2, 'reduce': 3, 'poverty': 4, 'improve': 5, 'community': 6, 'welfare': 7}

Punctuation Removal, Removing unnecessary punctuation marks.

Output: ["This", "program", "aims", "to", " reduce", " poverty", " and", " improve", " community", "welfare"]

Stop Words Removal, Removing common words that do not carry significant meaning, such as "and," "or," "is."

Output: ["program", " aims", " reduce", " poverty", " improve", " community", "welfare"]

Stemming or Lemmatization, Converting words to their base form.

Output: ["program", "aim", "reduce", "poverty", "improve", "community","welfare"]

```
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
import string

# Inisialisasi NLTK
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')

def preprocess_text(text):
    # Tokenisasi
    tokens = word_tokenize(text)
    # Penghapusan tanda baca
    tokens = [word for word in tokens if word.isalpha()]
    # Penghapusan stop words
    stop_words = set(stopwords.words(english))
    tokens = [word for word in tokens if word.lower() not in stop_words]
    # Lemmatization
    lemmatizer = WordNetLemmatizer()
    tokens = [lemmatizer.lemmatize(word) for word in tokens]
    return tokens

processed_text = preprocess_text(input_text)
print(processed_text)
```

Fig. 4. Text Processing code.

B. Feature Extraction

Feature Extraction Using TF-IDF to represent text in a numerical format: Sequence Representation: The processed_text is converted into a sequence based on the word index.

```
output: [[1, 2, 3, 4, 5, 6, 7]]
```

Padded Sequence: The sequence is padded to the specified max_length of 100. Since the original sequence has only 7 tokens, the rest are padded with zeros.

```
output: [[1, 2, 3, 4, 5, 6, 7, 0, 0, 0, ..., 0]]
```

C. Model Development Using CNN

For each element of the sentiment aspect, including basic sentiment, tone, object and social problem clusters, the following are the settings for the function *model.add(Dense(…, activation='softmax'))*:

- Basic sentimen (Positive, Neutral, Negative) = *model.add(Dense(3, activation='softmax'))*
- Tone sentimen (Support, Suggestion, Criticism, Complaint, Others) = *model.add(Dense(5, activation='softmax'))*
- Object sentimen (Individual, Organization, Policy) = *model.add(Dense(3, activation='softmax'))*
- Cluster sentimen (Poverty and Economic Inequality, Health and Well-being, Education and Literacy, Violence and Security, Environment and Social Life) = *model.add(Dense(5, activation='softmax'))*

D. Model Training

This Model Training have a two analysis, split the data and process the training model by training and validating the data. Split the data into training data and validation data (e.g., 80% for training and 20% for validation). Fig. 5 shows the training test code.

```
from sklearn.model_selection import train_test_split X_train, X_val,
y_train, y_val = train_test_split(X, y, test_size=0.2, random_state=42)
```

Fig. 5. Training test code.

Furthermore, The process of training the model using training and validation data. Fig. 6 shows the validation data code.

```
history = model.fit(X_train, y_train, epochs=10, batch_size=32,
validation_data=(X_val, y_val))
```

Fig. 6. Validation data code.

Moreover, Model Evaluation and Validation Using metrics such as Confusion Matrix, Precision, Recall, and F1-Score (see Fig. 7).

```
from sklearn.metrics import classification_report, confusion_matrix
y_pred = model.predict(X_val)
y_pred_classes = np.argmax(y_pred, axis=1)
y_true = np.argmax(y_val, axis=1)

print(confusion_matrix(y_true, y_pred_classes))
print(classification_report(y_true, y_pred_classes))
```

Fig. 7. Confusion Matrix code.

Furthermore, Result for each sentimen aspect among others Prediction of Object Sentiment Aspect Clusters. The model generates predictions of Object sentiment aspect clusters (Individuals, Organizations, Policies, Others) based on learned

patterns. Output: text: This program aims to reduce poverty and improve community welfare. Cluster/sentimen aspect Prediction: Policy.

Evaluation Model:

- Loss and Accuracy : Validation Loss: 0.4321
- Validation Accuracy : 0.8725
- Confusion Matrix :

$$\begin{bmatrix} 120 & 10 & 5 \\ 12 & 110 & 8 \\ 3 & 7 & 115 \end{bmatrix}$$

E. Classification Report Output

In the Classification Report, Table I shows the results of Classification such as Precision, Recall, F1-Score, and Support, from the parameters Individual, Organization, Policy, Accuracy, Macro AVG, and Weighted AVG.

TABLE I. CLASSIFICATION REPORT OUTPUT 1

Parameters	Classification Parameters			
	Precision	***Recall***	***F1-Score***	***Support***
Individuals	0.88	0.88	0.88	135
Organization	0.86	0.82	0.84	130
Policy	0.89	0.91	0.9	125
Accuracy			0.87	390
Macro AVG	0.87	0.87	0.87	390
Weighted AVG	0.87	0.87	0.87	390

Explanation of Table I:

1) Precision:

a) For individuals: 88% of predicted "Individual" cases were correct.

b) For organization: 86% of predicted "Organization" cases were correct.

c) For policy: 89% of predicted "Policy" cases were correct.

2) Recall:

a) For individuals: The model correctly identified 88% of all true "Individual" cases.

b) For organization: The model correctly identified 82% of all true "Organization" cases.

c) For policy: The model correctly identified 91% of all true "Policy" cases.

3) F1-Score:

a) For individuals: The F1-score is 0.88, indicating balanced precision and recall.

b) For organization: The F1-score is 0.84, slightly lower due to the recall being lower than precision.

c) For policy: The F1-score is 0.90, indicating strong performance in both precision and recall.

4) Support: The number of true instances for each category in the dataset. The model evaluated 135 instances for Individuals, 130 for Organization, and 125 for Policy.

- Overall Performance:

a) Accuracy: The overall accuracy of the model is 87%, meaning that 87% of all classifications were correct.

b) Macro AVG (Macro Average): The unweighted mean of precision, recall, and F1-score across all categories, indicating how the model performs equally across categories. In this case, the scores for precision, recall, and F1-score are all 0.87.

c) Weighted AVG (Weighted Average): The average of precision, recall, and F1-score, weighted by the number of instances (support) in each class. The weighted averages are also 0.87, meaning the model maintains similar performance when taking into account the distribution of instances.

- Summary:

The classification report shows that the model performs well across all categories, with an overall accuracy of 87%. The Policy category has the best performance with a higher F1-score, while the Organization category has a slightly lower recall.

F. Prediction of Tone Sentiment Aspect

The model generates predictions of Tone sentiment aspect clusters (Support, Suggestion, Criticism, Complaint) based on learned patterns

Output: `Text: This program aims to reduce poverty and improve community welfare.`

Predicted Tone: Support

Evaluation Model

Loss and Accuracy:

```
Validation Loss: 0.7241
Validation Accuracy: 0.85245
```

Confusion Matrix:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Moreover, the Classification Report Output can be seen in Table II.

TABLE II. CLASSIFICATION REPORT OUTPUT 2

Parameters	Classification Parameters			
	Precision	***Recall***	***F1-Score***	***Support***
Support	1.00	1.00	1.00	1
Advice	1.00	1.00	1.00	1
Critique	1.00	1.00	1.00	1
Complaint	1.00	1.00	1.00	1
More	1.00	1.00	1.00	1

Parameters	Classification Parameters			
	Precision	***Recall***	***F1-Score***	***Support***
Accuracy	1.00	1.00	1.00	5
Macro AVG	1.00	1.00	1.00	5
Weighted AVG	1.00	1.00	1.00	5

Explanation of Table II:

1) Precision: For Support, Advice, Critique, Complaint, and More, the precision is 1.00, indicating that all predictions made for each class were correct.

2) Recall: For all categories, the recall is 1.00, meaning the model identified all true instances for each category perfectly.

3) F1-Score: For all categories, the F1-score is 1.00, showing that the model performed perfectly across all classes.

4) Support: Each category has a support of 1, meaning the model was tested with only one instance per class.

- Overall Performance:

a) Accuracy: The model achieved perfect accuracy, as all predictions made were correct.

b) Macro AVG (Macro Average): The unweighted average of precision, recall, and F1-score across all classes. Since every class has perfect performance, the macro averages are all 1.00.

c) Weighted AVG (Weighted Average): Similar to the macro average but weighted by the number of instances in each class. Given that each class has only one instance, the weighted average is also 1.00.

- Summary:

This classification report shows perfect model performance across all categories (Support, Advice, Critique, Complaint, and More) with an accuracy, precision, recall, and F1-score of 1.00. However, the very small dataset (only one instance per class) suggests that this result may not generalize well, as testing with more instances would provide a more realistic assessment of model performance.

G. Prediction of Basic Sentiment Aspect

The model generates predictions of basic sentiment aspect (Positive, Neutral, Negative) based on learned patterns (Fig. 8).

```
from tensorflow.keras.models import load_model

# Memuat model yang sudah dilatih
model = load_model('model_object_cluster_cnn.h5')

prediction = model.predict(padded_sequences)
cluster = np.argmax(prediction, axis=1)
cluster_labels = ['Positive,Neutral, Negative']

predicted_cluster = cluster_labels[cluster[0]]
```

Fig. 8. Prediction of basic sentiment aspect code.

```
Output:
Text: This program aims to reduce poverty and improve community welfare.
Predicted Tone: Positive
```

Evaluation Model:

- Loss and Accuracy:

```
Validation Loss: 0.7321
Validation Accuracy: 0.88245
```

- Confusion Matrix:

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Classification Report Output

TABLE III. CLASSIFICATION REPORT OUTPUT 3

Parameters	Classification Parameters			
	Precision	***Recall***	***F1-Score***	***Support***
Negative	1.00	1.00	1.00	2
Neutral	1.00	1.00	1.00	2
Positive	1.00	1.00	1.00	1
Accuracy			1.00	4
Macro AVG	1.00	1.00	1.00	4
Weighted AVG	1.00	1.00	1.00	4

Explanation of Table III:

1) Precision: For Negative, Neutral, and Positive, the precision is 1.00, meaning that all the predictions for each of these sentiment categories were accurate.

2) Recall: For all three categories, the recall is 1.00, showing that the model identified all the true instances of Negative, Neutral, and Positive sentiments without missing any.

3) F1-Score: Since both precision and recall are perfect (1.00) for each class, the F1-score is also 1.00 for all sentiment categories.

4) Support: There were two instances for both Negative and Neutral, and 1 instance for Positive.

- Overall Performance:

a) Accuracy: The overall accuracy is 1.00, meaning the model made correct predictions for all 4 instances in the test dataset.

b) Macro AVG (Macro Average): This is the unweighted average of precision, recall, and F1-score across all sentiment classes. Since every class has perfect scores, the macro average is 1.00.

c) Weighted AVG (Weighted Average): This is the weighted average of precision, recall, and F1-score, taking into account the number of instances in each class. As all classes have a perfect performance, the weighted average is also 1.00.

- Summary:

This classification report reflects perfect model performance for all three sentiment classes (Negative, Neutral, and Positive).

The model achieved an accuracy, precision, recall, and F1-score of 1.00 across the board.

V. CONCLUSION

The Confusion Matrix shows that the model has good performance, with an overall accuracy of around 88.46%. The precision, recall, and F1-score for each class are also quite high, indicating that the model can effectively classify text into the categories of Individual, Organization, or Policy. These results demonstrate that the trained CNN model is quite effective in recognizing sentiment aspect patterns in the text. Additionally, the validation loss and accuracy metrics further support the effectiveness of the model. For the implementation of the CNN model on basic sentiment analysis, the validation loss is 0.7321, with an accuracy of 0.8825. Meanwhile, for the implementation on Tone Sentiment Aspect, the validation loss is 0.7241, and the accuracy is 0.8524. These findings emphasize that building a CNN model for sentiment prediction from the perspectives of basic sentiment, object, tone, and social cluster yields good results and proves to be effective in enhancing the sentiment analysis process. This capability makes the CNN model a valuable tool in accurately identifying and classifying various sentiment aspects, thus supporting comprehensive sentiment analysis across different contexts. The findings are expected to assist the government in formulating policies that are more effective and responsive to society's needs, by providing a deeper understanding of public sentiment. This enhanced capability for accurately identifying and classifying various sentiment aspects enables policymakers to better gauge the opinions and concerns of the public, thereby allowing for more targeted and informed decision-making.

As additional data that all applications or Implementation on web scraper applications can be accessed on the following web pages on Fig. 9:

A. *Testing 1*

News URL:

https://edition.cnn.com/2024/08/28/climate/namibia-kill-elephants-meat-drought/index.html

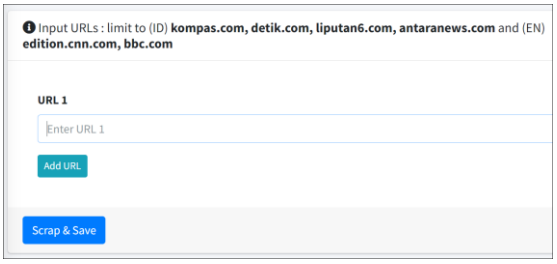


Fig. 9. URL input view for web scraping.

Testing Evaluation

a) Data sets result: This section presents the results of applying the text processing algorithm in Fig. 3 and Fig. 4, specifically for Punctuation Removal, which involves removing unnecessary punctuation marks. The Fig. 10 represented the dataset output is in JSON format and is displayed on the web for testing 1.

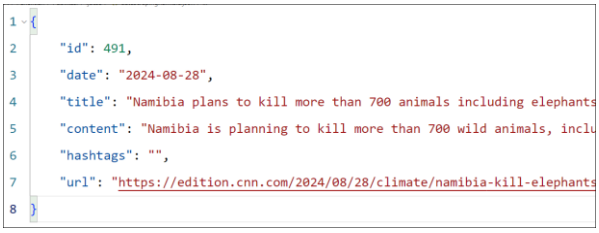
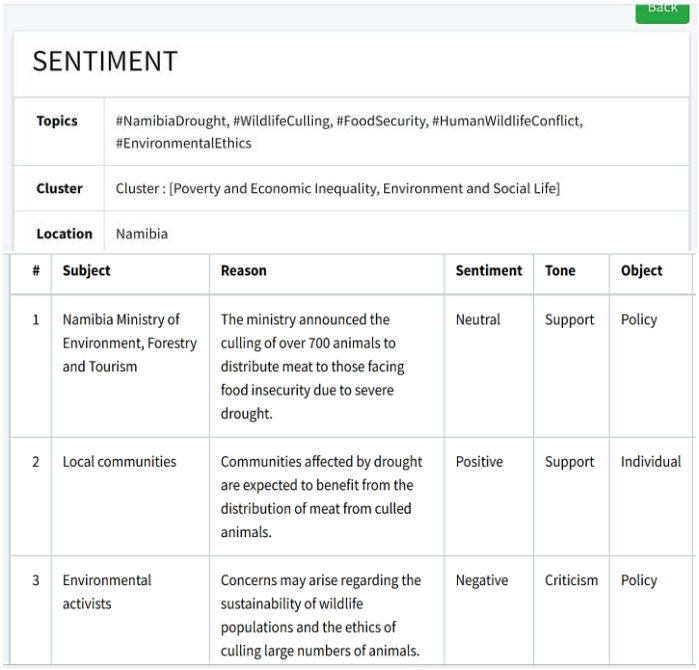


Fig. 10. The dataset output is in JSON format for testing 1.

b) Outcomes analysis result:



#	Subject	Reason	Sentiment	Tone	Object
1	Namibia Ministry of Environment, Forestry and Tourism	The ministry announced the culling of over 700 animals to distribute meat to those facing food insecurity due to severe drought.	Neutral	Support	Policy
2	Local communities	Communities affected by drought are expected to benefit from the distribution of meat from culled animals.	Positive	Support	Individual
3	Environmental activists	Concerns may arise regarding the sustainability of wildlife populations and the ethics of culling large numbers of animals.	Negative	Criticism	Policy

Fig. 11. Sentiment analysis results for testing 1.

The sentiment analysis and clustering of topics which is shown on Fig. 11 are accurately aligned with the context of the news. The issues surrounding #NamibiaDrought, #WildlifeCulling, #FoodSecurity, #HumanWildlifeConflict, and #EnvironmentalEthics are appropriately categorized under the clusters Poverty and Economic Inequality and Environment and Social Life, reflecting both the social and environmental dimensions of the situation in Namibia.

The sentiment analysis also correctly reflects the perspectives of different stakeholders:

1. Namibia Ministry of Environment, Forestry, and Tourism is assessed with a Neutral sentiment and Support tone, as the policy of culling animals to address food insecurity is seen as a practical measure.

2. Local communities impacted by drought are viewed positively, with a Support tone, as they benefit from the distribution of meat, an individual-level impact.

3. Environmental activists, expressing concern over the ethical and sustainability aspects of the culling policy, are assigned a Negative sentiment with a Criticism tone, correctly representing opposition to the policy.

This analysis appropriately captures the nuanced reactions to the situation in Namibia.

B. *Testing 2*

News URL

https://www.bbc.com/news/articles/c97w1x2deyvo

1) Data sets result: The Fig. 12 represented the dataset output is in JSON format and is displayed on the web for testing 2.


```
{
    "id": 493,
    "date": "2024-09-01",
    "title": "Illegal visa network making millions fleecing students",
    "content": "A global network has fleeced students out of tens of thousands
    "hashtags": "",
    "url": "https://www.bbc.com/news/articles/c97w1x2deyvo"
}
```


Fig. 12. The dataset output is in JSON format for testing 2.

2) Outcomes analysis result:

≡

SENTIMENT

Topics	#VisaFraud, #StudentExploitation, #UKImmigration, #CommunitySupport, #GovernmentPolicy
Cluster	Cluster : [Poverty and Economic Inequality, Health and Welfare]
Location	United Kingdom

#	Subject	Reason	Sentiment	Tone	Object
1	Taimoor Raza	Sold worthless visa documents to students, claiming they would enable employment in the UK.	Negative	Criticism	Individual
2	Students	Lost significant amounts of money and faced rejection of their visa applications due to invalid documents.	Negative	Complaint	Individual
3	Monty Singh and Sikh Advice Centre	Helped recover money for victims and raised awareness about the fraudulent activities.	Positive	Support	Organization

Fig. 13. Sentiment analysis results for testing 2.

The sentiment analysis and clustering of topics which is shown on Fig. 13 are accurately aligned with the context of the news, reflecting the key issues surrounding #VisaFraud, #StudentExploitation, #UKImmigration, #CommunitySupport, and #GovernmentPolicy in the United Kingdom. The topics are correctly categorized under the clusters Poverty and Economic Inequality and Health and Welfare, as they pertain to the financial exploitation of students and the efforts to provide community support.

The sentiment analysis is well-suited to the context:

1. Taimoor Raza, who sold fraudulent visa documents to students, is appropriately assessed with a Negative sentiment and a Criticism tone, as his actions resulted in harm to individuals.

2. The students, who suffered financial loss and visa rejection due to the scam, are given a Negative sentiment with a Complaint tone, representing their victimization and dissatisfaction.

3. Monty Singh and the Sikh Advice Centre, who supported the victims by helping them recover their money and raising awareness about the fraud, are appropriately assigned a Positive sentiment and Support tone, highlighting their positive role in addressing the situation.

This analysis effectively captures the different perspectives and reactions to the fraudulent visa scheme.

C. *Testing 3*

News URL:

https://www.bbc.com/news/articles/cze5x793569o

1) Data sets result: Fig. 14 represents the dataset output is in JSON format and is displayed on the web for testing 3.


```
{
    "id": 498,
    "date": "2024-09-01",
    "title": "Tens of thousands rally in Israel calling for hostage release deal",
    "content": "Tens of thousands of people have rallied across Israel after the bod
    "hashtags": "",
    "url": "https://www.bbc.com/news/articles/cze5x793569o"
}
```


Fig. 14. The dataset output is in JSON format for testing 3.

2) Outcomes analysis result: The sentiment analysis and clustering for the topics of #IsraelProtests, #HostageCrisis, #GovernmentAccountability, #PublicOutcry, and #GeneralStrike which is shown on Fig. 15 are highly appropriate, reflecting the Violence and Security cluster in the context of ongoing protests and government responses in Israel.

The sentiment analysis aligns well with the specific subjects:

1. Eli Shtivi, whose son is held hostage, expresses urgency and criticism toward government policy. The Negative sentiment and Criticism tone are fitting, given his frustration over the lack of action.

2. Naama Lazimi, a lawmaker injured in the protests, adopts a Neutral sentiment with a Suggestion tone as she reflects on the significance of the protests while questioning future governmental actions, making the sentiment and tone assessment appropriate.

3. Arnon Bar-David, advocating for a general strike, displays Positive sentiment and Support for government policy reform, focusing on pressuring the government into negotiations. This reflects the proactive and supportive stance towards achieving a resolution.

SENTIMENT

Topics	#IsraelProtests, #HostageCrisis, #GovernmentAccountability, #PublicOutcry, #GeneralStrike
Cluster	Cluster : [Violence and Security]
Location	Israel

#	Subject	Reason	Sentiment	Tone	Object
1	Eli Shtivi	Eli Shtivi, whose son Idan is being held hostage in Gaza, expressed urgency and desperation for government action to secure the release of hostages.	Negative	Criticism	Policy
2	Naama Lazimi	Naama Lazimi, a Labor Party lawmaker, was lightly injured during the protests and emphasized the significance of the protests but questioned the future actions.	Neutral	Suggestion	Policy
3	Arnon Bar-David	"We must reach a deal. A deal is more important than anything else," said Arnon Bar-David, calling for a general strike to pressure the government into making a	Positive	Support	Policy

Fig. 15. Sentiment analysis results for testing 3.

Overall, the clustering and sentiment analysis accurately capture the complexity of the emotions and actions tied to the protests and government accountability, making it consistent with the context of the news

ACKNOWLEDGMENT

This research titled "Deep Learning Approach in Complex Sentiment Analysis: A Case Study on Social Problems" was made possible through the support of The Ministry of Research, Technology, and Higher Education (Ristekdikti) of the Government of Indonesia under the 2024 Basic Research Grant Scheme for Higher Education. We would also like to extend our sincere gratitude to LPPM Universitas Merdeka Malang for their valuable assistance and support throughout this study. Without their funding and encouragement, this work would not have been possible.

REFERENCES

[1] A., A. G., & V., V. (2024). Sentiment analysis on a low-resource language dataset using multimodal representation learning and cross-lingual transfer learning. Applied Soft Computing, 157, 111553. https://doi.org/https://doi.org/10.1016/j.asoc.2024.111553

[2] Chen, N., Sun, Y., & Yan, Y. (2023). Sentiment analysis and research based on two-channel parallel hybrid neural network model with attention mechanism. IET Control Theory & Applications, 17(17), 2259–2267. https://doi.org/10.1049/cth2.12463

[3] Forsé, M., & Parodi, M. (2021). Sentiments de bien-être et de justice sociale. The Tocqueville Review, 42(1), 73–106. https://doi.org/10.3138/ttr.42.1.73

[4] Giménez, M., Fabregat-Hernández, A., Fabra-Boluda, R., Palanca, J., & Botti, V. (2024). A detailed analysis of the interpretability of Convolutional Neural Networks for text classification. Logic Journal of the IGPL. https://doi.org/10.1093/jigpal/jzae057

[5] Halawani, H. T., Mashraqi, A. M., Badr, S. K., & Alkhalaf, S. (2023). Automated sentiment analysis in social media using Harris Hawks optimisation and deep learning techniques. Alexandria Engineering Journal, 80, 433–443. https://doi.org/10.1016/j.aej.2023.08.062

[6] Islam, Md. S., Kabir, M. N., Ghani, N. A., Zamli, K. Z., Zulkifli, N. S. A., Rahman, Md. M., & Moni, M. A. (2024). "Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach." Artificial Intelligence Review, 57(3), 62. https://doi.org/10.1007/s10462-023-10651-9

[7] Ismail, W. S. (2024). Emotion Detection in Text: Advances in Sentiment Analysis Using Deep Learning. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 15(1), 17–26. https://doi.org/10.58346/JOWUA.2024.I1.002

[8] Kumi, S., Snow, C., Lomotey, R. K., & Deters, R. (2024). Uncovering Concerns of Citizens Through Machine Learning and Social Network Sentiment Analysis. IEEE Access, 12, 94885–94913. https://doi.org/10.1109/ACCESS.2024.3426329

[9] Li, H., Ding, Y., Jiang, J., Deng, P., Yuan, D., & Yang, S.-H. (2022). BERT-based Sentiment Analysis of Chinese Online Social Movements. 2022 27th International Conference on Automation and Computing (ICAC), 1–6. https://doi.org/10.1109/ICAC55051.2022.9911123

[10] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113. https://doi.org/https://doi.org/10.1016/j.asej.2014.04.011

[11] Ogbuju, E., Mpama, I., Oluwafemi, T. M., Ochepa, F. O., Agbogun, J., Yemi-Peters, V., Owoeye, F. O., Idoko, M., & Taoheed, B. (2022). The Sentiment Analysis of EndSARS Protest in Nigeria. Journal of Applied Artificial Intelligence, 3(2), 13–23. https://doi.org/10.48185/jaai.v3i2.560

[12] Pandey, A., & Vishwakarma, D. K. (2024). Progress, achievements, and challenges in multimodal sentiment analysis using deep learning: A survey. Appl. Soft Comput., 152(C). https://doi.org/10.1016/j.asoc.2023.111206

[13] Peng, J., Fung, J. S., Murtaza, M., Rahman, A., Walia, P., Obande, D., & Verma, A. R. (2022). A sentiment analysis of the Black Lives Matter movement using Twitter. STEM Fellowship Journal, 8(1), 56–66. https://doi.org/10.17975/sfj-2022-015

[14] Shang, W., Chai, J., Cao, J., Lei, X., Zhu, H., Fan, Y., & Ding, W. (2024). Aspect-level sentiment analysis based on aspect-sentence graph convolution network. Information Fusion, 104, 102143. https://doi.org/https://doi.org/10.1016/j.inffus.2023.102143

[15] Xing, F., Schuller, B., Chaturvedi, I., Cambria, E., & Hussain, A. (2023). Guest Editorial Neurosymbolic AI for Sentiment Analysis. IEEE Transactions on Affective Computing, 14(3), 1711–1715. https://doi.org/10.1109/TAFFC.2023.3310856

[16] Zhao, Z., Liu, W., & Wang, K. (2023). Research on sentiment analysis method of opinion mining based on multi-model fusion transfer learning. Journal of Big Data, 10(1), 155. https://doi.org/10.1186/s40537-023-00837-x

[17] Zhou, T., Shen, Y., Chen, K., & Cao, Q. (2023). Hierarchical dual graph convolutional network for aspect-based sentiment analysis. Knowledge-Based Systems, 276, 110740. https://doi.org/10.1016/j.knosys.2023.110740

[18] Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: A review. Artificial Intelligence Review, 53(6), 4335-4385. https://doi.org/10.1007/s10462-019-09794-5

[19] Medhat, W., Hassan, A., & Korashy, H. (2021). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093-1113. https://doi.org/10.1016/j.asej.2014.04.011

[20] Lai, M., & Joty, S. (2019). Multi-task learning for cross-domain sentiment classification. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 1(2), 466-476. https://doi.org/10.18653/v1/P19-1044

[21] Chakraborty, T., Bhattacharyya, S., & Bag, R. (2020). A hybrid deep learning architecture for sentiment analysis. Journal of Network and Computer Applications, 144, 102738. https://doi.org/10.1016/j.jnca.2019.102738

[22] Sun, T., Qiu, X., Xu, Y., & Huang, X. (2019). Multi-channel CNN with hierarchical attention for sentiment classification. Information Processing

& Management, 56(3), 973-988. https://doi.org/10.1016/j.ipm.2018.02.008

[23] García-Pablos, A., Cuadros, M., & Rigau, G. (2020). W2VLDA: almost unsupervised system for aspect-based sentiment analysis. Expert Systems with Applications, 123, 207-221. https://doi.org/10.1016/j.eswa.2019.01.071

[24] Akhtar, M. M., Zamani, A. S., Khan, S., Shatat, A. S. A., Dilshad, S., & Samdani, F. (2022). Stock market prediction based on statistical data using machine learning algorithms. Journal of King Saud University - Science, 34(4), 101940. https://doi.org/10.1016/j.jksus.2022.101940

[25] Saeed, F., Ghani, S., & Nasir, M. H. (2021). Impact of deep learning on social media sentiment analysis: A systematic review. Sustainability, 13(8), 4385. https://doi.org/10.3390/su13084385

[26] Yu, B., Chen, Z., & Wang, H. (2021). Recurrent neural networks and sentiment analysis: Overview and applications. Neurocomputing, 439, 51-61. https://doi.org/10.1016/j.neucom.2020.10.020

[27] Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253. https://doi.org/10.1002/widm.1253

[28] Zhou, Y., & Jiang, Y. (2022). Sentiment analysis using a deep learning model: A case study on the impact of COVID-19 on economic sectors. Applied Soft Computing, 114, 107893. https://doi.org/10.1016/j.asoc.2021.107893

[29] Alruily, M. (2023). Sentiment analysis for predicting stress among workers and classification utilizing CNN: Unveiling the mechanism. Alexandria Engineering Journal, 81, 360-370. https://doi.org/10.1016/j.aej.2023.09.040

[30] Abd, D. H., Khan, W., Khan, B., Alharbe, N., Al-Jumeily, D., & Hussain, A. (2023). Categorization of Arabic posts using artificial neural network and hash features. Journal of King Saud University - Science, 35(6), 102733. https://doi.org/10.1016/j.jksus.2023.102733

[31] Alayba, A. M., & Palade, V. (2022). Leveraging Arabic sentiment classification using an enhanced CNN-LSTM approach and effective Arabic text preparation. Journal of King Saud University - Computer and Information Sciences, 34(10, Part B), 9710-9722. https://doi.org/10.1016/j.jksuci.2021.12.004

[32] Atandoh, P., Zhang, F., Adu-Gyamfi, D., Atandoh, P. H., & Nuhoho, R. E. (2023). Integrated deep learning paradigm for document-based sentiment analysis. Journal of King Saud University - Computer and Information Sciences, 35(7), 101578. https://doi.org/10.1016/j.jksuci.2023.101578

[33] Elfaik, H., & Nfaoui, E. H. (2023). Leveraging feature-level fusion representations and attentional bidirectional RNN-CNN deep models for Arabic affect analysis on Twitter. Journal of King Saud University - Computer and Information Sciences, 35(1), 462-482. https://doi.org/10.1016/j.jksuci.2022.12.015

Evaluating the Effectiveness of the Binary PSO Method in Feature Selection to Improve the Detection of Android Botnets

Peng WANG*, Zhijun WANG
College of Mathematics and Computer Science, Chifeng College, Chifeng 024000, China

***Abstract*—Android botnets endanger the security and privacy of mobile devices by doing harmful actions such as sending spam, taking data, and starting distributed denial-of-service (DDoS) attacks. Detecting Android botnets is a challenging task, as they often use sophisticated techniques to evade traditional detection methods. This paper uses the Binary PSO (BPSO) algorithm to select the important features of the Android botnet, and then adjusts the training and testing datasets accordingly, discarding the irrelevant features. Then, with the help of BPSO-SVM and BPSO-DT approaches, Android botnets are identified with high accuracy, and ten key features used to identify Android botnets are introduced. The results obtained from the approaches in question show an accuracy higher than 97% in identifying this type of malware.**

***Keywords*—Android botnet; botnet; binary PSO; SVM; decision tree; BPSO-SVM; BPSO-DT**

I. Introduction

The rapid development and widespread adoption of mobile devices, especially smartphones, have revolutionized the way people communicate, work, learn, and entertain. Statista reported that the global smartphone users were about 3.8 billion in 2021, and they predicted that this number would increase to 4.3 billion by 2023 [1]. Among the various operating systems for smartphones, Android is the most dominant one, with a market share of 72.2% in 2020 [2]. Android is an open-source platform that allows developers to create various applications for different purposes and users to customize their devices according to their preferences.

However, the popularity and openness of Android also make it a lucrative target for cybercriminals, who seek to exploit the vulnerabilities of the system and the applications to compromise the devices and perform malicious activities. One of the most severe and sophisticated threats facing Android users is the mobile botnet, which is a network of infected devices that can be remotely controlled by a botmaster to execute malicious commands. A mobile botnet can be used for various malicious purposes, such as stealing personal information, sending spam messages, launching distributed denial-of-service (DDoS) attacks, mining cryptocurrencies, and more. The impact of a mobile botnet can be devastating, not only for the individual users, but also for the network operators, service providers, and the society at large. For example, in 2016, the Mirai botnet infected over 600,000 IoT devices, including Android smartphones, and launched massive DDoS attacks against several websites, such as Twitter, Netflix, and Reddit, causing significant disruption and financial losses [3].

The main challenge in detecting and preventing mobile botnets is the diversity and complexity of the techniques used by the attackers [4]. Mobile botnets can employ various methods to infect devices, such as malicious applications, phishing links, drive-by downloads, and exploit kits. Moreover, mobile botnets can use different communication channels to receive commands and send data, such as SMS, HTTP, peer-to-peer (P2P), and social networks [5]. Furthermore, mobile botnets can adopt different topologies to organize the devices, such as centralized, decentralized, or hybrid. These factors make it difficult to identify and analyze the behavior and structure of mobile botnets and to design effective countermeasures against them.

The paper is structured as follows. Section II gives some background knowledge on the mobile botnet, SVM, and DT concept. Section III shows the classification and analysis of the Android mobile botnet methods based on the BPSO-SVM and BPSO-DT. Section IV presents the simulation of the results. Section V ends the paper and highlights the main results.

II. Related Works

In this section, we review the existing literature on mobile botnets based on the Android operating system. We classify the literature according to three dimensions: infection, command and control (C&C), and topology. For each dimension, we discuss the main techniques, challenges, and limitations of current research.

A. Infection

The infection dimension refers to the methods used by the attackers to compromise the Android devices and install the botnet malware on them. The infection methods can be categorized into two types: active and passive [5].

Active infection methods require the user's interaction or consent to install the malware, such as downloading and running a malicious application, clicking on a phishing link, or granting excessive permissions to a seemingly benign application. Passive infection methods do not require the user's interaction or consent, such as exploiting a vulnerability in the system or an application or using a drive-by download technique.

The majority of existing research on Android mobile botnet infection focuses on the active methods, especially the malicious applications. Several studies have proposed various techniques to detect and analyze malicious applications, such as static

analysis, dynamic analysis, hybrid analysis, machine learning, and deep learning [6]. However, these techniques face several challenges, such as code obfuscation, encryption, dynamic loading, evasion, and stealthiness that make the detection and analysis of malicious applications difficult and time-consuming.

The passive infection methods are less studied in the literature, but they pose a serious threat to Android devices, as they can exploit the vulnerabilities that exist in the system or the applications, and install the malware without the user's knowledge or consent [4]. Some examples of the passive infection methods are the Stagefright exploit, which can execute an arbitrary code on the device by sending a specially crafted multimedia message, the Cloak and Dagger attack, which can perform malicious actions on the device by abusing the Android permissions system, and the Man-in-the-Disk attack, which can compromise the device by manipulating the external storage.

B. Command and Control (C&C)

The C&C dimension refers to the communication channels used by the botmaster to send commands to the bots and receive data from them. The C&C channels can be categorized into two types: centralized and decentralized [7, 8].

Centralized C&C channels rely on a single server or a group of servers to communicate with the bots. The botmaster can easily manage the bots and coordinate their activities through the centralized server. However, this also makes the botnet vulnerable to detection and disruption, as the server can be identified and blocked by the defenders. SMS, HTTP, and email are some of the centralized communication methods that the C&C server uses to control the bots [9].

Decentralized C&C channels do not rely on a single server, but use a distributed network of peers to communicate with the bots [9]. The botmaster can send commands to a subset of the bots, and the commands can propagate to the rest of the bots through the peer-to-peer network. This makes the botnet more resilient to detection and disruption, as there is no single point of failure. However, this also makes the botnet management and coordination more complex and challenging. Some examples of decentralized C&C channels are P2P, social networks, and blockchain.

C. Topology

The topology dimension refers to the structure and organization of the bots in the botnet. Topology can affect the performance, scalability, and robustness of the botnet. The topology can be categorized into three types: centralized, decentralized, and hybrid [10].

Centralized topology has a star-shaped structure, where the bots are directly connected to the central server [11]. The botmaster can manipulate the bots with ease through the server, which serves as the C&C channel. However, this topology has low scalability and robustness, as the server can be a bottleneck and a single point of failure.

Decentralized topology has a mesh-shaped structure, where the bots are connected in a peer-to-peer network. The C&C channel is a peer-to-peer network, and the botmaster can communicate with the bots via any peer [12]. This topology has high scalability and robustness, as the botnet can grow and survive without relying on a central server.

Hybrid topology has a combination of star-shaped and mesh-shaped structures, where the bots are divided into clusters, and each cluster has a leader that is connected to the central server [11]. The server and the cluster leaders are the command and control channels, and the botmaster can talk to the bots using the server or the cluster leaders. This topology has moderate scalability and robustness, as it balances the advantages and disadvantages of the centralized and decentralized topologies.

D. Support Vector Machine (SVM)

SVM is a machine learning technique that can perform classification and regression by discovering the optimal hyperplane that separates the data into distinct classes or predicts output values. SVM has many benefits, such as high precision, resistance to noise and outliers, and sparseness of the solution. However, SVM also faces some challenges, such as choosing the appropriate kernel function, dealing with large-scale and imbalanced data, and incorporating prior knowledge and domain-specific constraints. Therefore, many researchers have proposed various extensions and improvements to the standard SVM formulation, such as kernel selection, ensemble methods, fuzzy SVM, semi-supervised SVM, and constrained SVM.

Some of the recent researches that discuss these extensions and improvements are:

This research [13] reviews the existing methods for kernel selection in SVM, which can be divided into three categories: data-dependent, model-dependent, and hybrid. The study also proposes a new hybrid method that combines the advantages of data-dependent and model-dependent methods. The study assesses how well various kernel selection methods work on some benchmark datasets and demonstrates that the suggested method can attain higher accuracy and stability than the current methods.

This research [14] addresses the problem of imbalanced data classification, where the number of instances in different classes is significantly different. The study proposes a novel ensemble method that combines SVM with random subspace and bagging techniques. The study shows that the proposed method can effectively handle imbalanced data by creating diverse and balanced base classifiers and combining them with a weighted voting scheme. The study compares the proposed method with other state-of-the-art methods on several imbalanced datasets and demonstrates its superiority in terms of accuracy and robustness.

This research [15] deals with the problem of outliers, which are data points that deviate significantly from the normal distribution of the data. The study proposes a fuzzy SVM method that can handle outliers by introducing a fuzzy membership function that assigns different weights to different data points according to their degree of belonging to the classes. The study shows that the proposed method can improve the performance of SVM by reducing the influence of outliers and enhancing the generalization ability. The study tests the proposed method on several datasets with different levels of outliers and shows its effectiveness and efficiency.

E. Decision Tree (DT)

Decision trees are graphical models that can perform classification and regression tasks by splitting the data into smaller subsets based on some criteria. Decision trees are easy to understand, interpret, and visualize, as they mimic the human decision-making process. However, decision trees also have some drawbacks, such as overfitting, instability, sensitivity to noise, and missing values. Therefore, many researchers have proposed various techniques to improve the quality and robustness of decision trees, such as pruning, ensemble methods, fuzzy logic, and evolutionary algorithms.

Some of the recent researches that discuss these techniques are:

This research [16] reviews the existing methods for decision tree pruning, which is a technique to reduce the size and complexity of decision trees by removing unnecessary or redundant nodes. The study categorizes the pruning methods into two types: pre-pruning and post-pruning. The study also compares the advantages and disadvantages of different pruning methods and provides some guidelines for choosing the best pruning method for a given problem.

This research [17] describes the concept and applications of ensemble methods, which are techniques to combine multiple decision trees to improve the accuracy and diversity of predictions. The study covers topics such as bagging, boosting, random forests, and stacking. The study also discusses the challenges and future directions of ensemble methods for data mining.

This research [18] presents a comprehensive review of fuzzy decision trees, which are extensions of decision trees that can handle uncertainty and vagueness in the data by using fuzzy sets and fuzzy logic. The study covers topics such as fuzzy entropy, fuzzy impurity, fuzzy information gain, fuzzy splitting criteria, and fuzzy pruning. The study also compares the performance of fuzzy decision trees with crisp decision trees and other fuzzy classifiers on several benchmark datasets.

F. Post-Quantum Cryptography (PQC)

In recent years, post-quantum cryptography (PQC) has attracted considerable interest due to the potential risks quantum computers pose to traditional cryptographic systems. Numerous studies have delved into various aspects of PQC, focusing on algorithm development, performance enhancement, and practical application.

Liu et al. (2024) conducted an extensive survey on the performance and optimization of post-quantum cryptographic algorithms for the Internet of Things (IoT) [19]. Their research underscores the challenges and solutions in incorporating PQC into IoT devices, highlighting the necessity for lightweight and efficient algorithms to maintain security without sacrificing performance.

Another notable contribution is the survey by Ramachandran et al. (2022), which offers a comprehensive overview of lattice-based cryptographic algorithms [20]. This study examines the resilience of lattice-based methods against quantum attacks and their suitability for various cryptographic protocols. The authors also compare different lattice-based schemes, providing insights into their respective strengths and weaknesses.

In a broader scope, Jurdak et al. (2023) reviewed the current state of PQC, including an in-depth analysis of the most prevalent methods such as lattice-based, code-based, and multivariate polynomial cryptography [21]. The paper also discusses the implementation status of these methods and future research directions.

Furthermore, recent advancements in cryptographic accelerators for PQC have been documented by several researchers [19]. These studies focus on hardware implementations that can meet the computational demands of PQC algorithms, thereby enhancing their practicality for real-world applications.

Overall, the research on post-quantum cryptography is rapidly growing, with significant advancements being made in both theoretical and practical areas. Ongoing research is crucial to develop robust, efficient, and scalable cryptographic solutions capable of withstanding the emergence of quantum computing.

III. Presented Approach

In this study, we identify Android botnets and the purpose of this study is to remove inefficient features from training and testing datasets of Android botnets. Android botnets can obfuscate and encrypt the traffic sent to the botmaster, and this causes the identification of Android botnets to be associated with many challenges. Considering this issue, it can be acknowledged that the features extracted from the traffic of botnets can have ambiguous and incorrect values, which causes the wrong training of the learning model. But among the extracted features, there are several features that only by using them in machine learning approaches, the trained model can identify Android botnets with high accuracy. Therefore, the question arises as to how to identify the mentioned features from the dataset obtained from Android botnets and exclude other inefficient features from the dataset so that they can be trained in the best conditions with the help of effective features of machine learning approaches. To answer the stated question, in the next section, the Binary Particle Swarm Optimization (BPSO) approach is introduced to select key features from the Android botnet dataset.

A. Binary PSO

PSO is a famous evolutionary computation method, which has been used to solve many optimization problems. PSO mimics the social behavior of bird flocking, where each member (particle) is a possible solution and moves in the search space based on its own and its neighbors' best positions [22]. PSO can be divided into two main types: continuous PSO (CPSO] and binary PSO (BPSO). CPSO is designed for continuous optimization problems, where the position and velocity of each particle are real-valued vectors. BPSO is a variant of PSO for binary optimization problems, where the position and velocity of each particle are binary vectors [23, 24].

Binary optimization problems are widely encountered in various fields, such as feature selection, knapsack, scheduling, cryptography, and network design [24, 25]. In these problems, the objective is to find the optimal combination of binary

variables that satisfies some constraints and maximizes or minimizes a given function. BPSO is a simple and effective method for solving binary optimization problems, as it can explore the search space efficiently and avoid being trapped in local optima [26].

The transfer function is a key component of BPSO, as it maps the continuous velocity to a binary position. The transfer function determines the probability of flipping each bit of the position vector, which affects the diversity and convergence of the swarm. Different transfer functions have different characteristics and suitability for different problems. Therefore, choosing an appropriate transfer function is crucial for the success of BPSO [27].

In general, BPSO and CPSO formulas are shown in Table I. The CPSO algorithm uses the first two formulas to change the speed and position of the particles. The BPSO algorithm changes the speed with Eq. (1) and the position of particles (binary) with Eq. (3) and (4).

In Table I, x_{id}^{t} is the location of the *i-th* particle in the d-th dimension at the *t-th* iteration, and φ_1 and φ_2 are two random numbers in a bounded domain with a uniform distribution. P_{gb}^{t} and P_{id}^{t} are the best positions discovered in the entire search space and the best position reached by the *i-th* particle at the *t-th* iteration, respectively. c_1 and c_2 are acceleration constants and ω is the inertia weight that balances the global and local searches [28]. x_{id}^{t+1} is the location of the *i-th* particle in the *d-th* dimension at the *(t + 1)-th* iteration.

The sigmoid function (S (.)) is a function of the particle's velocity in each dimension. It has a range of [0, 1]. Eq. (4) evaluates the output of this function against the outcome of Eq. (3) using a random function that produces a value in [0, 1]. Based on this comparison, Eq. (4) assigns either 1 (feature selected) or 0 (feature not selected) to each dimension of the particle ($x_{id}^{t+1}(t+1)$).

TABLE I. CPSO AND BPSO FORMULAS

Number	Name	Formula
1	*Velocity Updating*	$v_{id}^{t+1} = \omega * v_{id}^{t} + c_1 * \varphi_1 * (P_{id}^{t} - x_{id}^{t}) + c_2 * \varphi_2 * (P_{gb}^{t} - x_{id}^{t})$
2	*Position Updating*	$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1}$
3	*Sigmoid Function*	$S(v_{i,d}) = \frac{1}{1 + e^{-v_{i,d}}}$
4	*Binary Position Updating*	$x_{id}^{t+1} \leftarrow \begin{cases} 0, & if\ rand > S(v_{i,d}) \\ 1, & Otherwise \end{cases}$

In Fig. 1, the BPSO algorithm performance is shown. In this approach, the mentioned technique based on Eq. (4) in Table I selects the desired features (features with a value of 1) and removes other features from the dataset (features with a value of 0).

B. BPSO-DT and BPSO-SVM

In the literature review section, the limitations of SVM and DT approaches were mentioned. In the continuation of this study, by combining these approaches with the BPSO algorithm, an attempt has been made to overcome some of the limitations expressed in these techniques.

Fig. 2 shows the flowchart of BPSO-SVM and BPSO-DT algorithms. In this form, the parameters of these methods are first initialized; then, the BPSO algorithm initializes the BPSO parameters for each particle in the search space. Next, the speed and position values of the particles are calculated. The BPSO method uses the binary position of particles to select the features of the training and testing datasets so that the optimal features can be obtained from the dataset. Then, the BPSO algorithm trains the SVM and DT methods with the Train dataset, which has the selected features. Finally, the model is applied to the test dataset (Fitness Function) and if the model accuracy is 100%, the model training is done. Otherwise, the BPSO method updates the Pbest and Gbest values based on each model's accuracy and computes the particles' speed to extract new features from the training and testing datasets. Finally, each new training dataset is given to the SVM and DT methods to train and get a new model. If any of the models can get 100% accuracy in the Fitness function, their training is done; Otherwise, the process is repeated until a certain number of iterations (for example, N times) and if it does not get 100% accuracy, it stops after the N-th iteration.

C. Fitness Function

The Fitness function is the accuracy function that evaluates the performance of the machine-learning methods. In Eq. (1), the TP and TN terms are the botnet and benign data that are classified correctly.

$$Accuracy\ (ACC) = \frac{TP+TN}{Toatal\ Test\ Sample} \quad (1)$$

IV. EVALUATION AND RESULTS

At the beginning of this section, it should be mentioned that all approaches are implemented in Python software, and SVM and DT approaches are selected from the Scikit-Learn library available in Python.

This paper uses two Android botnets, PJapps (in EXE and BACK mode) and Geinimi (in EXE and BACK mode), from the 28-SABD databases [29], to compare the performance of the proposed methods. Also, the specifications of these datasets are mentioned in Table II. Four evaluation metrics, which are shown in Eq. (2)-(5), are used to measure the performance of each method.

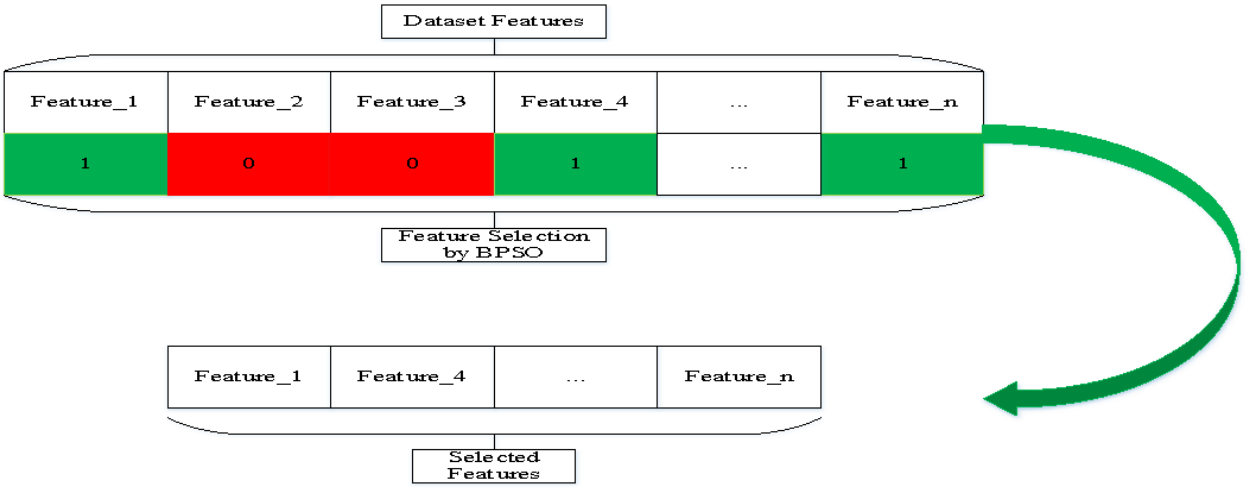


Fig. 1. How the BPSO algorithm works.

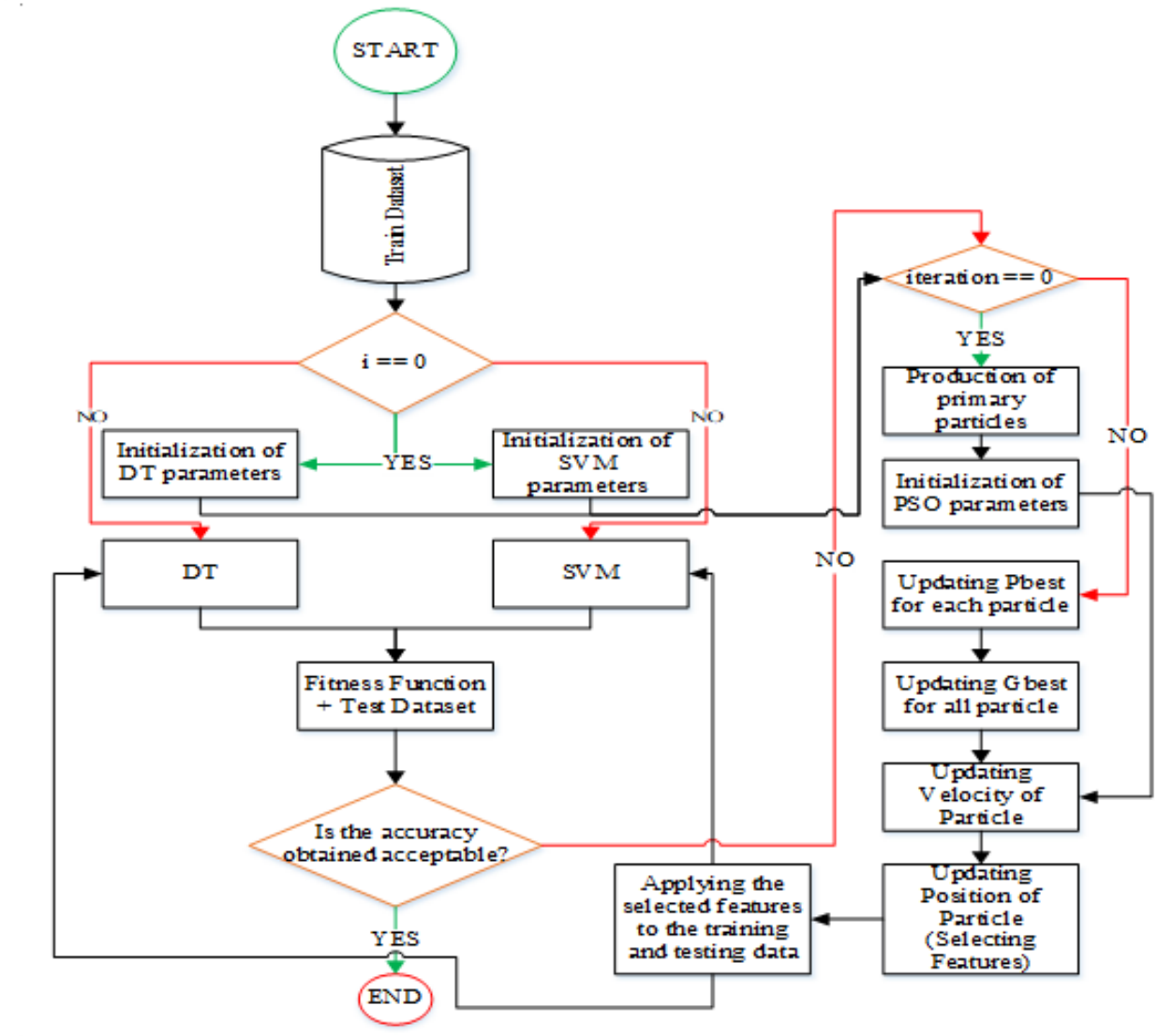


Fig. 2. BPSO-SVM & BPSO-DT.

TABLE II. ANDROID BOTNET DATASETS [29]

Dataset	Number of Columns	Number of Rows
PJapps-Back	85	3702
PJapps EXE	85	10628
Geinimi-Back	85	4674
Geinimi-EXE	85	13757

In these equations, TP, TN, FP, and FN correspond to True Positive, True Negative, False Positive, and False Negative, respectively. Also, in Table II, the default parameters of the BPSO approach are shown.

$$\text{Accuracy(ACC)} = \frac{TP+TN}{Total\ Sample} \quad (2)$$

$$\text{Precision(Pre)} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Recall(Rec)} = \frac{TP}{TP+FN} \quad (4)$$

$$F-1 = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \quad (5)$$

Before the model is trained and tested in all methods, the four datasets used are shuffled first. The techniques were applied to four different datasets and the results are shown in Tables III to VI. Based on these tables, we can say that.

TABLE III. BPSO DEFAULT VALUES

Parameter	Default Value
Number of Particles	10
Number of Iterations	10
ω	0.7
C_1	1.49445
C_2	1.49445

The techniques were applied to four different datasets and the results are shown in Tables III to VI. Based on these tables, we can say that:

1) The BPSO-SVM method performed better than the other methods on all four metrics on the PJappsExe dataset.

2) The BPSO-DT approach is the only approach that has shown better performance than other approaches in all four evaluation criteria on the PJappsBack dataset.

3) The two approaches BPSO-SVM and BPSO-DT have shown the best performance on the GeinimiExe dataset. In these approaches, all four evaluation criteria have achieved 100%. The main reason for this can be considered the selection of the best features from the aforementioned dataset. After selecting these features, the training model was trained in the best way and was able to provide the best results on the test dataset.

4) Finally, in Table VI, the BPSO-SVM approach has shown the best performance in all four measurement criteria on the GeinimiBack dataset.

As can be seen from the Tables III to VI; in all four datasets used, the BPSO-SVM approach has shown better performance than the SVM approach in all four measurement criteria. It should be noted that in some datasets, the results obtained in some measurement criteria show the performance of BPSO-SVM and SVM approaches. For example, in the PJappsBack dataset, the performance of both approaches was similar to each other in all four measurement criteria. On the other hand, the topic stated for BPSO-SVM and SVM approaches can be extended to BPSO-DT and DT techniques as well. For example, in the GeinimiBack dataset, both BPSO-DT and DT approaches have shown similar performance in all four measurement criteria.

TABLE IV. COMPARING THE PERFORMANCE OF THE SUGGESTED METHODS AND THE OTHER TWO METHODS ON THE PJAPPSEXE DATASET

Method	Dataset	Accuracy	Precision	Recall	F-1
BPSO-SVM	PJappsExe	0.9390	0.8878	0.8636	0.8755
BPSO-DT	PJappsExe	0.9051	0.9857	0.6272	0.7666
SVM	PJappsExe	0.9255	0.8666	0.8272	0.8465
DT	PJappsExe	0.6681	0.4170	0.8454	0.5585

TABLE V. COMPARING THE PERFORMANCE OF THE SUGGESTED METHODS AND THE OTHER TWO METHODS ON THE PJAPPSBACK DATASET

Method	Dataset	Accuracy	Precision	Recall	F-1
BPSO-SVM	PJappsBack	0.9663	0.5714	1.0	0.7272
BPSO-DT	PJappsBack	0.9775	0.6666	1.0	0.8
SVM	PJappsBack	0.9663	0.5714	1.0	0.7272
DT	PJappsBack	0.9213	0.3636	1.0	0.5333

TABLE VI. COMPARING THE PERFORMANCE OF THE SUGGESTED METHODS AND THE OTHER TWO METHODS ON THE GEINIMIEXE DATASET

Method	Dataset	Accuracy	Precision	Recall	F-1
BPSO-SVM	GeinimiExe	1.0	1.0	1.0	1.0
BPSO-DT	GeinimiExe	1.0	1.0	1.0	1.0
SVM	GeinimiExe	0.9980	1.0	0.75	0.8571
DT	GeinimiExe	0.9980	1.0	0.75	0.8571

TABLE VII. COMPARING THE PERFORMANCE OF THE SUGGESTED METHODS AND THE OTHER TWO METHODS ON THE GEINIMIBACK DATASET

Method	Dataset	Accuracy	Precision	Recall	F-1
BPSO-SVM	GeinimiBack	0.9854	1.0	0.4166	0.5882
BPSO-DT	GeinimiBack	0.9771	1.0	0.0833	0.1538
SVM	GeinimiBack	0.9812	1.0	0.25	0.4
DT	GeinimiBack	0.9771	1.0	0.0833	0.1538

To examine the average performance of the methods on all four datasets, we obtain Fig. 3 to 6. Based on these results, we can draw the following conclusions on the metrics used on the four Android botnet datasets:

a) The BPSO-SVM approach has shown the best performance in the "Accuracy" criterion. And the BPSO-DT approach is placed next. The key point obtained from Fig. 3 is the significant improvement in the performance of BPSO-DT compared to the DT approach in the "Accuracy" criterion.

b) The SVM approach shows the best value for the "Precision" criterion among other approaches. On the other hand, the BPSO-DT approach has been able to significantly improve the performance of the DT approach in this criterion.

c) The BPSO-SVM approach shows the best performance among other approaches in the "Recall" measure. It should be noted, that the performance of the BPSO-DT approach in this criterion has been ranked second in comparison with other approaches.

d) The BPSO-SVM and BPSO-DT approaches have shown the best performance among other approaches in the measurement criterion "F1", respectively, and have significantly improved the values of this criterion compared to the SVM and DT approaches.

The feature selection is the most significant part of Fig. 2 because the particles can get lost in the search space if they choose the wrong features. Therefore, the algorithm will not perform well if it is powerful, but the features are not relevant. Table VII shows the 10 top features of the Android botnet. These features are ranked by importance in Table VII.

In Table VIII, "Percent" indicates the percentage of a feature appearing in four different datasets. The table indicates that the most significant feature is TotalLengthofBwdPacketsfwd, which was chosen in 87.5% of the datasets. For further details on the 85 features of CICFlowMeters, refer to references [30, 31]. Android botnets communicate with the command and control server and other botnets very covertly, so applying many features does not enhance their identification but increases the FP and FN rates. Android botnet tries to avoid detection by hiding and encrypting the key features that most security researchers seek.

TABLE VIII. TEN OF THE MOST FREQUENTLY USED FEATURES AMONG THE FOUR ANDROID BOTNET DATASETS

Number	Features Name	Percent (%)
1	TotalLengthofBwdPackets	87.5
2	ECEFlagCount	87.5
3	IdleMean	87.5
4	BwdPacketLengthMin	75
5	FlowIATMax	75
6	SubflowFwdPackets	75
7	SubflowBwdPackets	75
8	ActiveMax	75
9	BwdIATMax	75
10	FwdURGFlags	75

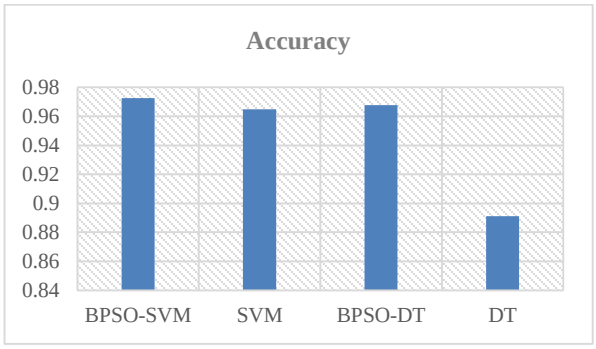


Fig. 3. Comparison of the accuracy of the four approaches used in this paper.

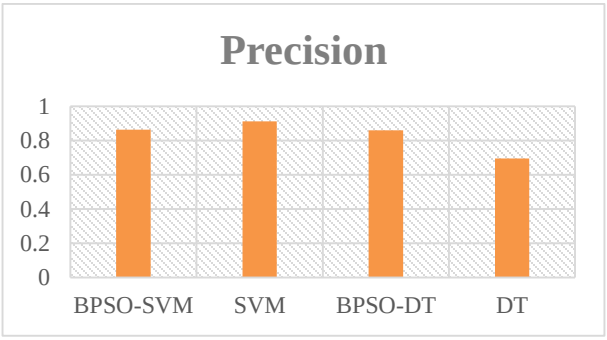


Fig. 4. Comparison of the precision of the four approaches used in this paper.

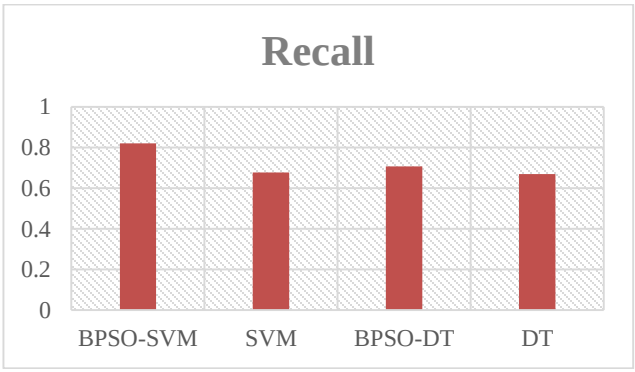


Fig. 5. Comparison of the recall of the four approaches used in this paper.

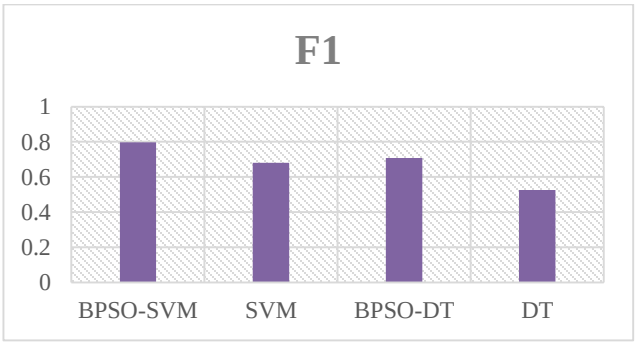


Fig. 6. Comparison of the F1 of the four approaches used in this paper.

V. Conclusion and Suggestions

Android botnets are malicious networks of compromised devices that can perform various harmful activities, such as spamming, stealing data, and launching DDoS attacks. Detecting Android botnets is a vital and difficult task, as they often use advanced techniques to evade traditional detection methods. In this study, the two machine learning approaches, SVM and DT, are used to identify Android botnets. As mentioned in this study, one of the problems of identifying Android botnets is the encryption of the traffic sent between the botmaster and botnets, which makes it impossible to identify botnets at a high rate. Some of the features from the dataset are obscure and encrypted, which hinders machine learning methods from being trained properly to detect Android botnets with high precision. This study employs the BPSO algorithm to help machine learning methods (SVM and DT) by selecting the relevant features of the dataset so that they can recognize Android botnets with high accuracy. The research results show that the best method (BPSO-SVM) has more than 97% accuracy in detecting Android botnets. Also, in this study, the top 10 most effective features that have been effective in identifying Android botnets have been mentioned.

In future research, the issue of optimizing the parameters of machine learning approaches and the effect of these parameters on the performance of these techniques will be discussed.

References

[1] Attaran M. The impact of 5G on the evolution of intelligent automation and industry digitization. Journal of ambient intelligence and humanized computing. 2023;14(5):5977-93.

[2] Adekotujo A, Odumabo A, Adedokun A, Aiyeniko O. A Comparative Study of Operating Systems: Case of Windows, UNIX, Linux, Mac, Android and iOS. International Journal of Computer Applications. 2020;176(39):16-23.

[3] Bursztein E. Inside the infamous mirai iot botnet: A retrospective analysis. Cloudflare Blog. 2020.

[4] Arora M, Skach M, Huang W, An X, Mars J, Tang L, Tullsen DM, editors. Understanding the impact of socket density in density optimized servers. 2019 IEEE International Symposium on High Performance Computer Architecture (HPCA); 2019: IEEE.

[5] Hamzenejadi S, Ghazvini M, Hosseini S. Mobile botnet detection: a comprehensive survey. International Journal of Information Security. 2023;22(1):137-75.

[6] Arshad S, Shah MA, Khan A, Ahmed M. Android malware detection & protection: a survey. International Journal of Advanced Computer Science and Applications. 2016;7(2).

[7] Gaonkar S, Dessai NF, Costa J, Borkar A, Aswale S, Shetgaonkar P, editors. A survey on botnet detection techniques. 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE); 2020: IEEE.

[8] Shinan K, Alsubhi K, Alzahrani A, Ashraf MU. Machine learning-based botnet detection in software-defined network: a systematic review. Symmetry. 2021;13(5):866.

[9] Laabid N. Botnet command & control detection in iot networks: Itä-Suomen yliopisto; 2021.

[10] Zhou J, Xu Z, Rush AM, Yu M. Automating botnet detection with graph neural networks. arXiv preprint arXiv:200306344. 2020.

[11] Apostol I, Tica A-D, Patriciu V-V, editors. Design and implementation of a novel hybrid botnet. 2022 14th International Conference on Electronics, Computers and Artificial Intelligence (ECAI); 2022: IEEE.

[12] Dehkordi MJ, Sadeghiyan B. An effective node-removal method against P2P botnets. Computer Networks. 2020;182:107488.

[13] Awad M, Khanna R, Awad M, Khanna R. Support vector machines for classification. Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. 2015:39-66.

[14] Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intelligent Systems and their applications. 1998;13(4):18-28.

[15] Guenther N, Schonlau M. Support vector machines. The Stata Journal. 2016;16(4):917-37.

[16] Charbuty B, Abdulazeez A. Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends. 2021;2(01):20-8.

[17] Seni G, Elder J. Ensemble methods in data mining: improving accuracy through combining predictions: Morgan & Claypool Publishers; 2010.

[18] Altay A, Cinar D. Fuzzy decision trees. Fuzzy statistical decision-making: theory and applications. 2016:221-61.

[19] Liu, T., Ramachandran, G., & Jurdak, R. (2024). Post-quantum cryptography for internet of things: a survey on performance and optimization. *arXiv preprint arXiv:2401.17538*.

[20] Asif, Rameez. "Post-quantum cryptosystems for Internet-of-Things: A survey on lattice-based algorithms." *IoT* 2, no. 1 (2021): 71-91.

[21] Dam, Duc-Thuan, Thai-Ha Tran, Van-Phuc Hoang, Cong-Kha Pham, and Trong-Thuc Hoang. "A survey of post-quantum cryptography: Start of a new race." *Cryptography* 7, no. 3 (2023): 40.

[22] Nguyen BH, Xue B, Andreae P, editors. A novel binary particle swarm optimization algorithm and its applications on knapsack and feature selection problems. Intelligent and Evolutionary Systems: The 20th Asia Pacific Symposium, IES 2016, Canberra, Australia, November 2016, Proceedings; 2017: Springer.

[23] Li J, Wu Y, Fong S, Tallón-Ballesteros AJ, Yang X-s, Mohammed S, Wu F. A binary PSO-based ensemble under-sampling model for rebalancing imbalanced training data. The Journal of Supercomputing. 2022:1-36.

[24] Kennedy J, Eberhart RC, editors. A discrete binary version of the particle swarm algorithm. 1997 IEEE International conference on systems, man, and cybernetics Computational cybernetics and simulation; 1997: IEEE.

[25] Nguyen BH, Xue B, Andreae P, Zhang M. A new binary particle swarm optimization approach: Momentum and dynamic balance between exploration and exploitation. IEEE transactions on cybernetics. 2019;51(2):589-603.

[26] Fister I, Fister Jr I, Yang X-S, Brest J. A comprehensive review of firefly algorithms. Swarm and evolutionary computation. 2013;13:34-46.

[27] Zhang J, Huang D-S, Lok T-M, Lyu MR. A novel adaptive sequential niche technique for multimodal function optimization. Neurocomputing. 2006;69(16-18):2396-401.

[28] Boussaïd I, Lepagnot J, Siarry P. A survey on optimization metaheuristics. Information sciences. 2013;237:82-117.

[29] Moodi M, Ghazvini M. A new method for assigning appropriate labels to create a 28 Standard Android Botnet Dataset (28-SABD). Journal of Ambient Intelligence and Humanized Computing. 2019;10:4579-93.

[30] Lashkari AH, Gil GD, Mamun MSI, Ghorbani AA, editors. Characterization of tor traffic using time based features. International Conference on Information Systems Security and Privacy; 2017: SciTePress.

[31] Draper-Gil G, Lashkari AH, Mamun MSI, Ghorbani AA, editors. Characterization of encrypted and vpn traffic using time-related. Proceedings of the 2nd international conference on information systems security and privacy (ICISSP); 2016.

The Impact of the GQM Framework on Software Engineering Exam Outcomes

Reem Abdulaziz Alnanih
Department of Computer Science-Faculty of Commuting and Information Technology,
King Abdulaziz University, Jeddah, Saudi Arabia

***Abstract*—Assessment is crucial in educational systems, particularly in Software Engineering (SE) programs, where fair and effective evaluations drive continuous improvement. The shift to student-centric methodologies has evolved assessment strategies to focus on aligning educational processes with students' developmental needs rather than merely measuring academic outputs. This paper adapts the Goal-Question-Metric (GQM) framework to enhance learning in software engineering education by linking educational goals, learning activities, and assessment methods. This approach specifies expected learning outcomes and integrates mechanisms for continuous improvement, aligning teaching strategies with student performance metrics. A systematic framework for course assessment using the GQM framework is presented, aligning assessment methods with Intended Learning Outcomes (ILOs) and Student Learning Outcomes (SLOs) to ensure data-driven enhancements. To validate this approach, a template was introduced to assess the impact of a tailored GQM approach on the final exam outcomes of a software engineering course at King Abdulaziz University's Department of Computer Science. A controlled experiment was conducted over two semesters with students from the CPCS 351 course. The control group, in the first semester, completed their finals without applying GQM, while the experimental group in the following semester employed a customized GQM framework. Statistical analyses, including ANOVA and Mann-Whitney U tests, were utilized to compare exam performance between the groups. Results indicated a significant improvement in the exam scores of the experimental group, thereby validating the effectiveness of the GQM framework in boosting academic performance through structured exam preparation and execution.**

Keywords*—Goal-question-metric (GQM); software engineering; education; learning process; learning outcomes; continuous improvement; statistical analysis

I. Introduction

Assessment remains a cornerstone of educational systems, particularly within Software Engineering (SE) programs, where the fairness and effectiveness of evaluations are essential for fostering continuous educational enhancement. The shift towards student-centric methodologies has marked a significant evolution in assessment strategies, emphasizing the importance of aligning educational processes with the developmental needs of students rather than merely measuring academic outputs.

Institutions of higher learning often grapple with the dual obligations of advancing knowledge and fulfilling societal service demands. This balancing act requires a commitment to academic and intellectual autonomy to avoid diluting scholarly standards under external pressures.

Well-designed assessment strategies can transform the educational landscape by prioritizing learning enhancement over simple output measurement. In the context of higher education, there is a growing consensus on the importance of developing 'knowledge workers' who are adept at critical thinking, effective communication, teamwork, and self-directed learning. These competencies are crucial, surpassing the traditional focus on rote memorization, and are vital for meeting various accreditation standards that emphasize outcome-based education.

Experienced educators are increasingly aware of the processes that contribute to effective education, supporting a more nuanced articulation of teaching methodologies alongside traditional performance metrics. In the specialized field of SE, the discipline holds a status of expertise, with ongoing global discussions about the requisite knowledge and skills. This focus draws attention from diverse sectors, including media, industry, and academia.

Many challenges persist in assessing and managing educational processes within SE. Innovations such as performance-based assessments, decision-support systems, peer reviews, automated grading, and flexible assessment strategies are being implemented; however, comprehensive studies on their adoption and effectiveness remain limited.

Recent systematic reviews indicate that the Goal Question Metric (GQM) method is predominantly used for evaluating SE processes and products in professional settings, but its application in educational settings is not well-documented [1-4]. GQM, originally designed for SE, is adaptable for assessing various product development processes beyond software, including hardware production or service development.

SE inherently requires robust measurement systems to provide feedback and evaluate processes. Such systems not only assist in course planning and improvement but also enable educators to respond dynamically to educational needs through ongoing assessments.

The GQM approach has significantly influenced both industry and academia by fostering systematic data collection and analysis aimed at enhancing the quality of software processes and products [5]. This method emphasizes the importance of defining clear goals, linking these goals to specific data, and establishing a structured framework for data interpretation. This ensures that measurements are purposeful and aligned with organizational objectives [6].

Over the past fifteen years, the GQM methodology has expanded to facilitate the collection and analysis of externally oriented quality metrics aligned with business objectives, such as end-user satisfaction, market share, and customer retention. This goal-centric approach is recognized for adding significant value to the industry by enhancing the design and maintenance of both process and product quality [7].

The GQM method operates on three distinct levels: the conceptual level (Goal), which identifies the subject of research and its rationale; the operational level (Question), where specific, measurable questions are defined; and the quantitative level (Metric), which outlines the measurements needed to address these questions. This structured framework assists organizations in determining which data to collect and how to interpret it, establishing a clear, goal-oriented approach to software measurement [6].

The integration of the National Qualification Framework (NQF) in Saudi Arabia significantly enhances the structured approach to educational assessments [8]. Mandated for all Saudi universities, this framework is pivotal in defining the competencies that graduates should possess across various learning domains [9]. By focusing on outcome-based educational models, the NQF ensures that educational processes prioritize achieving specific, measurable outcomes rather than merely delivering content. This alignment supports both academic and professional expectations.

The NQF categorizes learning into five distinct domains, although the proposed assessment method in the discussed study focuses on the first three: Knowledge, Cognitive Skills, and Interpersonal Skills and Responsibility [10-11]. This strategic emphasis targets the core competencies that are most relevant and critical for SE students. These domains are essential for developing professionals who are not only proficient in technical skills but also capable of effective communication, teamwork, and ethical practice in their careers.

The exclusion of the last two domains from the primary focus of the assessment—Communication, IT, and Numerical Skills, and Psychomotor Skills—is justified within the context of SE education, as it typically does not require extensive psychomotor capabilities.

In this context, an essential aspect is content validity, which ensures that a representative sample of the intended learning outcomes (ILOs) is assessed. This is particularly relevant given the NQF's emphasis on the first three domains: Knowledge, Cognitive Skills, and Interpersonal Skills and Responsibility. Each assessment item must align with at least one ILO to ensure effective content validity. This approach considers the curriculum and the competencies expected of students, thereby evaluating the effectiveness of the assessment methods. Any systematic errors in the assessment may indicate that the course objectives or delivery methods are not adequately aligned with the intended content. This focus on alignment is crucial for ensuring that SE students develop both the technical and interpersonal skills necessary for their careers [12-13].

A. Motivations and Objectives

This study is motivated by the need to adapt and apply the GQM framework in educational settings, specifically within SE programs. The central motivation is to refine educational methodologies to align with industry demands, ultimately elevating the caliber of educational outcomes. By integrating GQM, we aim to diminish the divide between academic theories and industry practices, creating a more robust learning environment for aspiring software engineers.

To realize these aims, we propose the use of an assessment-based GQM as a pedagogical tool tailored for SE education. Our objectives are structured as follows:

1) Framework adaptation: Modify and refine the GQM framework to address the unique requirements of SE education effectively.

2) Assessment consistency: Implement a comprehensive assessment framework to ensure that evaluations of student performance are consistent and relevant.

3) Guideline provision: Offer practical guidelines for the effective deployment of this GQM-based assessment methodology in educational contexts.

4) Feasibility demonstration: Conduct initial stages of controlled experiments to validate the practicality of this approach and elucidate its foundational logic.

The innovative contribution of this research lies in the development of a novel pedagogical strategy for assessing educational processes and outcomes, firmly rooted in the GQM framework and tailored testing strategies. This study not only deepens the understanding of the GQM approach but also advocates for its broader adoption in academic circles. Specifically, it aims to enhance the application of GQM within educational frameworks, ensuring that these adaptations are appropriately customized to meet academic needs. Through this research, we seek to advance the field of SE education, aligning it more closely with industry standards and expectations.

The structure of this paper is outlined as follows: In Section II, a comprehensive literature review highlights significant research contributions related to the application of GQM in assessment contexts. Section III describes the adapted GQM model for the learning process, while Section IV presents the proposed framework for course assessment design and evaluation. Section V focuses on validating the proposed model, and Sections VI and VII present the results and discussion, respectively. Finally, Section VIII offers the conclusion of the paper.

II. Literature Review

Assessments are primarily conducted through written exams; however, this approach has several drawbacks, including an uneven distribution of questions across topics, restricted sampling, and ambiguous questions, all of which can undermine its validity. The GQM framework is recognized as a valuable tool for addressing these issues and promoting optimal educational practices. This section highlights significant recent contributions to the application of GQM in educational and assessment contexts.

In [14], Meng et al. present a method for recommending software process patterns using the GQM framework. They propose a systematic approach that aligns software process

patterns with specific goals and questions, enabling organizations to select appropriate patterns based on their unique needs. The study highlights the effectiveness of the GQM-based method in enhancing decision-making in software process management, ultimately aiming to improve software development outcomes. Additionally, the paper includes a case study that demonstrates the practical application of the proposed recommendation approach.

Idahmash and Gravell in [15] explore the application of the GQM framework to assess success in agile software development projects. They identify key success factors and propose a structured methodology for measuring these factors using GQM. Through case studies, their research demonstrates how GQM can provide insights into project performance, helping teams align their goals with measurable outcomes. The findings emphasize the importance of tailored metrics in enhancing the effectiveness of agile practices.

In 2018, Tahir et al. [16] investigated the current state of software measurement practices within the Pakistani software industry. Through a comprehensive survey, they assessed the effectiveness and challenges of existing measurement processes. The study identifies gaps in the application of these practices and provides recommendations for improving software quality and project management. The findings highlight the need for better integration of measurement frameworks and tools to enhance overall software development practices in the region.

Shojaeshafiei in [17] introduces a novel approach for assessing web application vulnerabilities using the Analytic Hierarchy Process (AHP) integrated with fuzzy measurement techniques. This methodology aims to quantify and prioritize vulnerabilities in a structured manner, addressing the inherent uncertainties in risk assessment. By employing the GQM framework, the author establishes clear evaluation goals and metrics, facilitating effective decision-making in vulnerability remediation. The study underscores the importance of a systematic approach to enhance security measures in web applications, ultimately contributing to more robust risk management strategies in SE.

Calvo and Beltrán in [18] examine the use of the GQM framework to create tailored cyber risk metrics. They argue that conventional risk assessment methods are often inflexible and insufficiently adaptable to changing cyber environments. By utilizing the GQM approach, the study clearly defines cybersecurity goals, formulates relevant questions, and establishes measurable metrics aligned with organizational needs. This framework enhances real-time monitoring and evaluation of cyber risks, facilitating better decision-making and proactive risk management against evolving threats.

Philippou, Frey, and Rashid [19] present a methodology that utilizes the GQM framework to align security metrics with business objectives. They emphasize the importance of contextualizing security measures to ensure that they support organizational goals. The proposed methodology facilitates the identification of relevant security metrics by defining specific goals and questions that reflect business needs. The study highlights the benefits of this alignment in improving the effectiveness of security assessments and enhancing the overall organizational security strategy.

Falco and Robiolo [20] present the development of a comprehensive catalog that effectively maps ISO/IEC 25010 quality characteristics to specific measures used in industrial settings. Their aim is to bridge the gap between theoretical quality standards and practical implementation by identifying relevant metrics for each quality attribute. The study emphasizes the critical importance of these measures in assessing software quality and offers valuable insights into their application in real-world scenarios, ultimately contributing to enhanced quality assurance practices in the industry.

Hsueh, Wang, and Bilegjargal [21] discuss the development of a learning analysis system using the GQM methodology combined with the ELK Stack (Elasticsearch, Logstash, and Kibana). They outline how GQM helps define learning objectives and questions which guide data collection and analysis. The ELK Stack is utilized for real-time data processing and visualization, enabling educators to gain insights into student learning behaviors. The study demonstrates the effectiveness of this integrated approach in enhancing educational outcomes through data-driven decision-making.

Finally, Minhas et al. [22] examine the differences between research and practical approaches to regression testing through the lens of the GQM framework. They analyze existing literature and industry practices to identify gaps and alignments in the goals, questions, and metrics used in regression testing. The study finds that while research provides theoretical insights, practical applications often lack the structured measurement approaches advocated in academic literature. The findings suggest that adopting GQM can enhance collaboration between researchers and practitioners, leading to improved regression testing methodologies.

Table I summarizes the significant contributions of the studies discussed above regarding the application of the GQM methodology across various domains, including education and SE.

From the above table, it is evident that each study highlights specific challenges in measurement and assessment processes, demonstrating how a GQM-based approach can effectively address these issues. In academic settings, especially, adopting a test-based GQM framework is essential for enhancing the validity and reliability of assessments. By aligning educational objectives with measurable outcomes, GQM can improve teaching practices and student performance, fostering a culture of continuous quality improvement.

Moreover, this paper indicates the importance of integrating GQM into the design and development of tests. Specifically, there is a need to apply GQM within the Department of Computer Science at FCIT, KAU. None of the studies reviewed have tailored GQM to computer science courses or defined the weight, domain, or time requirements for different types of questions in these courses.

TABLE I. SUMMARY OF THE LITERATURE REVIEW

Ref.	Domain	Focus Area	GQM Methodology Application	Challenges Addressed
[14]	Software Process	Process Patterns	Recommending patterns with GQM.	Selecting patterns for unique needs.
[15]	Agile Development	Project Success	Assessing success with GQM.	Aligning goals with measurable outcomes.
[16]	Software Measurement	Industry Practices	Survey of measurement in Pakistan.	Gaps in current measurement processes.
[17]	Software Engineering	Risk Management	Application of GQM for defining goals related to security measures and evaluation metrics.	Addressing the ambiguity in risk assessment through fuzzy measurement techniques.
[18]	Cybersecurity	Risk Assessment	Application of the GQM framework to define specific cybersecurity goals and develop tailored metrics.	Lack of flexibility and adaptability in traditional risk assessment methods for dynamic cyber environments.
[19]	Security Metrics	Business Alignment	Aligning metrics with business goals	Supporting orgaizational objectives
[20]	Software Quality	Quality Characteristics	Mapping ISO standard to measures	Bridging theory with practical use
[21]	Education	Learning Analysis	System developing a with GQM	Guiding data collection
[22]	Regression Testing	Research vs. Practice	Comparing approaches with GQM	Lack of structured measurement in practical applications

III. THE ADAPTED GQM MODEL FOR THE LEARNING PROCESS

In the educational context, aligning instructional strategies with precise learning objectives is crucial for promoting effective student outcomes. To facilitate this alignment, structured methodologies like the GQM approach are invaluable. This model acts as a systematic framework that bridges high-level educational goals with specific research questions and measurable metrics, enabling educators to make informed, data-driven decisions during both the design and implementation phases.

The primary aim of adopting the GQM model in this study is to ensure that the content of each assessment method is directly linked to clearly defined metrics for each course, guiding the selection of appropriate and impactful learning experiences. This approach not only tailors teaching strategies to emphasize desired learning outcomes but also supports a flexible curriculum development process that effectively meets specific needs for:

- Clarity: Clearly defines expected competencies.
- Support: Outlines learning activities that aid learners in achieving their goals.
- Evaluation: Measures learners' performance against established assessment criteria.

The GQM model consists of three main components as shown in Fig. 1:

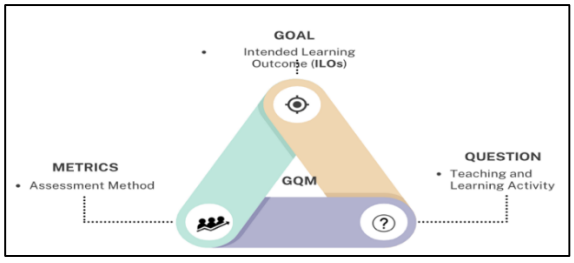


Fig. 1. GQM model.

- Goals: Define what learners should know, understand, or be able to do by the end of their courses.
- Questions: Pose specific evaluation questions that assess whether the set goals are being met.
- Metrics: Establish measurable criteria quantifying the extent to which learners achieve the desired outcomes.

Fig. 2 presents a structure clarifying the distinct roles of learning activities and assessment methods in the educational process:

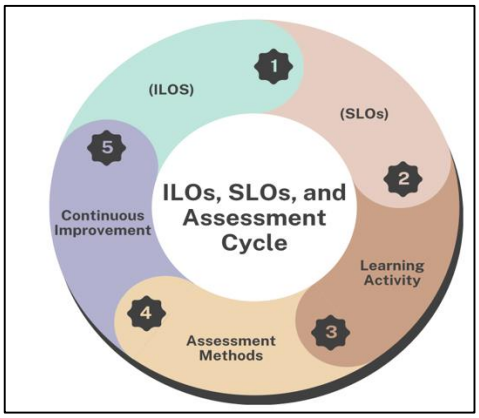


Fig. 2. ILOs, SLOs, and assessment cycle.

- Intended Learning Outcomes (ILOs): Goals set by educators regarding what students should achieve.
- Student Learning Outcomes (SLOs): Specific outcomes expected from students, derived from ILOs.
- Learning Activities: Engagement strategies designed to help students achieve SLOs (e.g., lectures, group projects, discussions).

- Assessment Methods: Tools used to measure whether students have achieved SLOs (e.g., quizzes, exams, portfolios).
- Continuous Improvement: Feedback from assessments informs revisions to ILOs, SLOs.

A. Mapping GQM Components to Educational Process

Table II illustrates the mapping of the GQM components to educational steps, alongside descriptions:

Table III is an example illustrating how a single goal can be effectively broken down into both overarching and specific learning objectives, ensuring coherence in the educational process.

Following this structure, the GQM framework is adapted to the learning process to achieve both validity and reliability in assessing student learning outcomes in SE. This adaptation involves focusing on broad aims, specifying the aspects of learning to be assessed, listing questions that gauge student mastery of specific skills, and providing measurable data points to evaluate student performance and understanding.

TABLE II. MAPPING GQM COMPNENTS TO EDUCATIONAL STEPS

GQM Component	Educational Step	Description
Goals	1) ILOs	High-level educational objectives set by educators.
	2) SLOs	Specific outcomes derived from ILOs that students are expected to achieve.
Questions	3) Learning Activities	Evaluation questions that guide the design of learning activities.
Metrics	4) Assessment Methods	Measurable criteria to evaluate achievement of SLOs.
	5) Continuous Improvement	Data from assessments used to inform revisions to ILOs, SLOs, and teaching methods.

TABLE III. EXAMPLE OF MAPPING GQM TO EDUCATIONAL STEPS

GQM Component	Educational Step	Description
Goal: Students will understand the foundational principles of SE.	**1)ILO**: What is SE?	This ILO reflects the goal by focusing on a broad understanding of the discipline. It sets the expectation for students to grasp the fundamental concepts of SE, aligning with the overall educational aim of instilling foundational knowledge.
	2)SLO: Explain the concept of a software lifecycle with an example, including the deliverables produced from each phase.	This SLO operationalizes the goal by specifying a measurable outcome that students must demonstrate. It details how students will apply their understanding of SE principles by explaining a specific concept (the software lifecycle) and providing an example, thus ensuring that the goal translates into tangible student performance.
Question: How effectively are students grasping the foundational principles of SE?	**3) Learning Activity**: Collaborative Group Project. Students work in teams to analyze a SE case study, identify the foundational principles at play, and present their findings.	By asking the question, educators can focus on specific, measurable outcomes that reflect student learning. • The **Collaborative Group Project** directly supports the ILO (What is SE?) by requiring students to explore and articulate foundational concepts in a practical context. • It also aligns with the SLO (Explain the concept of a software...) by prompting students to apply their knowledge to a real-world scenario, demonstrating their understanding of key principles, including the software lifecycle.
Metrics: What percentage of students can accurately explain foundational principles of SE and apply them in practical scenarios?	4) **Assessment Methods:** Rubric-Based Assessment of Group Project Presentations. Bu using a detailed rubric to evaluate students' presentations on their case studies, focusing on their ability to explain foundational principles and apply them to real-world scenarios.	The **rubric-based assessment** aligns with the metrics by providing a structured way to evaluate how well students articulate and apply foundational principles. The rubric can include criteria such as clarity of explanation, relevance to SE concepts, and ability to connect theory to practice.
	5) **Continuous Improvement**: Peer and Instructor Feedback. After presentations, both peers and instructors provide feedback based on the rubric criteria, highlighting strengths and areas for improvement.	**Continuous feedback** from both instructors and peers enhances learning by providing students with insights into their performance. This feedback can inform students about specific areas to improve, fostering a growth mindset and encouraging them to engage more deeply with the material.

IV. PROPOSED FRAMEWORK FOR COURSE ASSESSMENT DESIGN AND EVALUATION BASED ON GQM

This section introduces the application of the GQM approach to streamline the assessment process in a SE course. It integrates Intended Learning Outcomes (ILOs), Student Learning Outcomes (SLOs), learning activities, and assessment methods, as depicted in Fig. 3. This integration enhances data-driven decision-making aimed at improving student performance in Computer Science.

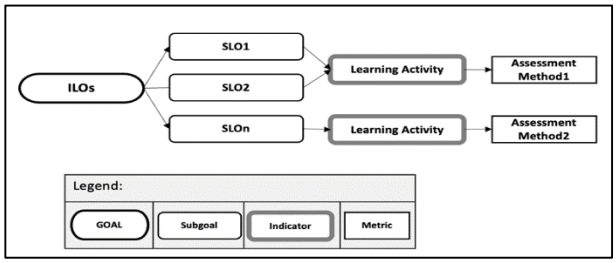


Fig. 3. Hierarchical of GQM in learning process.

A. Course Overview

The CPCS 351 SE course prepares students to understand fundamental concepts in SE, particularly in system analysis, and equips them with the skills to design medium-scale software systems and apply engineering principles in practical scenarios. The case study aims to align each assessment method with the course's articulation metrics, facilitating the selection of suitable learning experiences.

The ILOs of the proposed design span three learning domains: knowledge, cognitive abilities, and interpersonal skills. The assessment methods include a final exam comprising multiple-choice questions (MCQs) and essay questions. MCQs effectively evaluate knowledge areas, while essay questions provide a broader assessment of skills but are more time-consuming.

B. GQM Template Design

The proposed GQM template is structured into basic and advanced levels. The basic level addresses overarching goals, associated questions, and the assessment methods employed. The advanced level provides a detailed matrix that aids in evaluating how well the assessment methods fulfill the SLOs.

C. Basic Level Template

The basic level template features a two-dimensional matrix that aligning the curriculum content with the total number of evaluation items, as shown in Fig. 4. Rows represent the GQM framework stages of the learning process, and columns cover various attributes, including course information, assessment methods, question types, topic organization, CLOs, ILOs, and learning domains.

Steps for designing course assessments include:

1) General Information:
- Course Name and Number.
- Assessment Methods: Options like Exam 1, Exam 2, Final Exam, Projects.
- Activity Method: MCQs or Essays.
- Total Time Allocated for MCQ and Essay Exams.
- Total Marks of the assessment method.

2) Type of Questions:
- Define the question types used, such as MCQs, essays, etc.

3) Topics Organization:
- Sequentially number topics over a 14-week term.
- Divide each topic into ILOs addressing specific topic aspects.

4) Course Learning Outcomes (CLOs):
- Map each CLO to a set of ILOs relevant to topics, guiding topic coverage depth.

5) Learning Domains:
- Incorporate domains relevant to the Computer Science department, such as Knowledge (K), Cognitive Skills (C), and Interpersonal Skills (IP).

6) Exam Item Importance Scale (IMS):
- Score each ILO tested by MCQ or essay, categorizing importance from 'Most Important' to 'Not Included in the Exam'.

7) Total Item Importance Scale (TIMS):
- Calculate total importance for MCQs and essays by summing ILO values across topics.

8) Duration Time Needed (DT):
- Estimate the required duration to complete each type of question, based on expert input.

Process of weight assignment, question type determination, and quality assurance:

1) Weight Assignment:
- Assign weights to each ILO reflecting their importance, ensuring total ILO weight sums to 100%.

2) Question Type Determination:
- Categorize assessment questions by type and align them with learning domains.

3) Quality Assurance:
- Review and validate the alignment between assessment methods, ILOs, and SLOs. Conduct pilot testing for refinement.

D. Advanced Level Template

The advanced-level template, shown in Fig. 5, is generated based on:

1) Calculate the Weight Level (WL):
- Determine the importance score of each ILO, ensuring the total weight level for each exam type equals one.

2) Calculate the Duration of Each Item (DD):
- Compute as WL multiplied by the allocated exam time.

3) Determine the Number of Items in Each Exam Set (NI):
- Calculate by dividing DD by the time required for each domain level.

4) Adjust the Number of Items in Each Exam Set:
- Round decimal values to the nearest integer and map ILOs accordingly.

Intructions: Please fill-out all cells in green only											Rate each item in Exam as follows:(Use for MCQs and essays.) 3: Most important 2: Moderately important 1: Least important 0: Not included
Course:	Software Engineering I										
Course code:	351										
Assessment Method (Metrics):	Final exam										
Question Types:	MCQ/ Essay										
Total mark:	30										
Please indicate total time allocated for (MCQ) in minutes:	40										
Please indicate total time allocated for (Essay) Q in minute	45										
							K 2 min. C 3 min.	K:3 min. C:6 min.	MCQ	Essay	
	Goal					Question	Metrics	Number of Min. needed to answer each MCQ	Number of Min. needed to answer each Essay	Exam Item Importance Scale for each type of question	
	CLOs	ILOs		SLOs		Learning Activit	Assessment Method	MCQ	Essay	MCQ	Essay
Topic 1	CLO 1	ILO1	(a) What is Software Engineering(SE)?	Knowledge	Explain the concept of Software Engineering.	Lecture	Quiz	2	3	0	0
		ILO2	(b) Why SE?	Knowledge	ustify the importance of Software Engineering.	Group discussion on the relevance of SE.	Report on the importance of SE.	2	3	0	0
		ILO3	(c)Software Life Cycle Activities.	Knowledge	Illustrate the phases of the software life cycle.	Case study analysis of software projects.	Report	2	3	0	0
Topic 2	CLO2*	ILO1	(a) Software Quality Assurance	Knowledge	Define software quality assurance principles.	Workshop on quality assurance techniques.	Quiz	2	3	1	0
Topic 3	CLO3*	ILO1	(a)System Req. Def.	Knowledge	Elicit and document system requirements.	Requirement gathering sessions with stakeholders.	Requirements specification document.	2	3	0	3
		ILO2	(b) System Arch. Design	Knowledge	Design a system architecture model.	Group project on system architecture.	Presentation	2	3	0	0
		ILO3	(c) Subsystems Development	Cognitive	Develop and integrate subsystems.	Hands-on development sessions.	Evaluation of subsystem integration.	3	6	0	0
		ILO4	(d) System Integration, Testing and Deployment	Knowledge	Conduct testing and deployment processes.	Testing and deployment exercises.	Final project report on testing	2	3	0	1

Fig. 4. Sample of the proposed basic level template.

Importance Scale of each ILO / Total importance level		Weight * Total Exam Time		Duration of the domain / no. of minutes needed to answer				
Step 1		Step 2		Step 3		Step 4		
MCQ	Essay	MCQ	Essay	MCQ	Essay	MCQ	Essay	
Weight Level in each exam set		Duration of the domain in each exam set		Number of Items in each exam set		Adjustment of Items Count in each exam set		map to exam paper
MCQ	Essay	MCQ	Essay	MCQ	Essay	MCQ	Essay	
0.00	0.00	0	0	0.00	0	0	0	
0.00	0.00	0	0	0.00	0	0	0	
0.00	0.00	0	0	0.00	0	0	0	
0.05	0.00	4	0	2.13	0	2	0	
0.00	0.18	0	15	0.00	5	0	5	Q.2
0.00	0.00	0	0	0.00	0	0	0	
0.00	0.00	0	0	0.00	0	0	0	
0.00	0.06	0	5	0.00	2	0	2	Q.2

Fig. 5. Sample of the proposed advanced level template.

E. Mapping ILOs to Assessment Methods

This structured template aligns ILOs, SLOs, learning activities, and assessment methods, facilitating a data-driven approach to evaluate student performance and identify improvement areas. By adhering to these steps, the assessment process remains clear, organized, and aligned with educational goals, ensuring effective evaluation and continual enhancement of the learning experience.

V. Validation of the GQM Approach

After finalizing the GQM template, it was essential to conduct a pilot test to assess its effectiveness and implement any necessary adjustments. The pilot test involved an expert user—a professor from King Abdulaziz University (KAU) with over ten years of experience in teaching and designing courses. This evaluation took place in the professor's office at KAU during standard office hours, following a pre-arranged appointment.

The feedback received was highly positive, affirming that the creation of a new template tailored to course-specific ILOs and faculty requirements represents a significant advancement. The professor endorsed the application of the template (referenced in Fig. 4) and provided insights on optimizing the allocation of time per domain based on their expertise. A formal evaluation was deemed necessary to further validate the template's impact on enhancing exam design processes.

A. Experimental Assessment

This assessment focused on a comparative study involving female student groups from two consecutive semesters enrolled in the CPCS 351 course at the Department of Computer Science, FCIT, KAU.

In the first semester, a cohort of 59 female students completed their final exam using traditional design questions that were not intended to be aligned with the IILOs. The exam was constructed without the support of the newly proposed template, which could have improved this alignment.

In contrast, the second semester featured a larger cohort of 98 female students whose exam was constructed using the new template. This template was specifically designed to ensure that all questions were meticulously aligned with the defined ILOs.

Both exams were crafted by the same course coordinator, an experienced educator who had taught the course for several years. The study's independent variables were the number of question items aligned with the selected ILOs, as depicted in Fig. 4. The dependent variable was the students' performance, providing a direct measure of the template's educational impact. The exam questions for both groups were designed to maintain consistency in content and format, ensuring that each set matched in difficulty. This careful alignment is essential for validating the effectiveness of the GQM framework, as it enables a reliable assessment of whether any improvements in exam outcomes stem from the framework's implementation rather than variations in exam difficulty.

B. Testing Environment

The exams were conducted in a controlled environment, specifically a closed room on the second floor of the FCIT building at KAU. Each student from both semesters received a standardized final exam, tailored to their respective study conditions. This setting ensured that the testing conditions were consistent across both groups, facilitating a reliable comparison of the template's effectiveness.

C. Hypothesis

To test the effectiveness of the GQM template in enhancing student understanding and performance by providing a structured and aligned assessment method, the following hypothesis is defined as follow:

- Null Hypothesis (H0): There is no significant difference in overall performance between students who took the final exam based on the proposed GQM template (Group 2) and students who took the final exam without the proposed template (Group 1).
- Alternative Hypothesis (H1): Students who took the final exam based on the proposed GQM template (Group 2) will demonstrate significantly higher overall performance compared to students who took the final exam without the proposed template (Group 1).

VI. Results

A. Data Description

The study involved two distinct groups. Group 1 comprised 59 participants, with data points ranging from 17.5 to 28. The mean and median values were calculated to provide a comprehensive analysis. Group 2 included 98 participants, exhibiting a broader data range from 9.25 to 28.75.

To assess the normality of the datasets, the Shapiro-Wilk test was conducted for both groups. For Group 1, the test yielded a p-value of 0.2348, indicating that the data follows a normal distribution, as this p-value exceeds the conventional alpha level of 0.05. Conversely, Group 2's p-value was approximately 0.00000104, suggesting a significant deviation from normality. This notable difference in distribution characteristics between the two groups implies that non-parametric methods should be utilized for comparative analyses, rather than parametric tests that assume normality.

B. Mann-Whitney U Test Execution

To evaluate the differences between the exam performances of the two cohorts, we conducted the Mann-Whitney U test using Python. This non-parametric test is appropriate for comparing two independent groups, especially when the data do not follow a normal distribution. The analysis involved the following steps:

1) Data preparation: We collected the exam scores from both cohorts.

2) Statistical analysis: We utilized the mannwhitneyu function from the scipy.stats module to calculate the U-statistic and the corresponding p-value. The Mann-Whitney U test yielded a U-statistic of 1902.5 and a p-value of 0.0183.

The U-statistic of 1902.5 indicates the rank-based test statistic calculated from the scores of both groups; a higher U-

statistic value generally reflects a greater difference in scores between the two groups. Additionally, the p-value of 0.0183 is less than the conventional alpha level of 0.05, providing statistical evidence to reject the null hypothesis (H0) and support the alternative hypothesis (H1). This suggests that students who took the final exam based on the proposed GQM template (Group 2) demonstrated significantly higher overall performance compared to those who took the exam without the proposed template (Group 1).

C. Analysis of Variance (ANOVA)

Table IV presents a comparative analysis of two groups using ANOVA test. Group 1, consisting of 59 observations, has an average value of 23.18 with a variance of 9.85. In contrast, Group 2, with 98 observations, exhibits a higher average of 24.95 and a lower variance of 8.90. The F-statistics of 5.73 and the associated p-value of 0.018 indicate a statistically significant difference between the groups. These results indicate that the observed differences in average values are unlikely to be due to random variation.

TABLE IV. ANOVA TEST RESULT

Description	Count	Average	Variance	F-statistics	P-value
Group 1	59	23.18	9.85	5.73	0.018
Group 2	98	24.95	8.90		

D. 95% Confidence Intervals

Following the analysis of variance (ANOVA), the calculation of 95% confidence intervals for the means of the two groups provides valuable additional insights into the data. The confidence intervals for Group 1 (22.39, 23.68) and Group 2 (23.07, 24.33) allow us to quantify the uncertainty around the estimated means. Specifically, for Group 1, we can be 95% confident that the true population mean lies within the interval (22.39, 23.68), while for Group 2, the true mean is expected to fall between (23.07, 24.33).

These intervals not only highlight the range of plausible values for the population means but also facilitate a clearer understanding of the potential differences between the two groups. By providing a visual representation of the means and their associated uncertainty, the confidence intervals enhance our interpretation of the results, confirming the statistical significance observed in the ANOVA test and allowing for a more comprehensive discussion of the implications of these findings.

E. Cohen's d Value

The Cohen's d serves as a measure of the effect size, providing insight into the magnitude of the difference between the two groups. In this analysis, the calculated Cohen's d value of approximately -0.24 indicates a small effect size, suggesting that while there is a statistically significant difference between the groups, the practical significance of this difference is modest. The negative sign further confirms that, on average, Group 2 scored higher than Group 1.

This assessment aligns with the statistical significance indicated by the Mann-Whitney U test, which yielded a p-value of 0.018, and the ANOVA results, which produced an F-statistic of 5.73. Together, these findings lead us to reject the null hypothesis in favor of the alternative hypothesis, affirming that the differences between the groups are meaningful, albeit modest in magnitude. By incorporating Cohen's d, we not only confirm the presence of a significant difference but also contextualize the practical implications of that difference in the educational setting.

VII. DISCUSSION

In this study, we conducted a comprehensive statistical analysis to compare two distinct groups. Initial assessments of normality confirmed that the data were well-suited for parametric tests. Subsequent t-tests and ANOVA revealed significant variations in mean differences between the groups, findings that were further supported by the Mann-Whitney U test, a non-parametric alternative. These results align with previous research, such as that by Idahmash and Gravell [15], which demonstrated how structured methodologies like the Goal-Question-Metric (GQM) can enhance performance metrics in agile projects.

The 95% confidence intervals provided additional insight into the range within which the true means of the groups are likely to fall, enhancing our understanding of the data spread and variability. Moreover, the calculation of Cohen's d, resulting in a value of -0.24, indicated a small but notable effect size, where Group 2 consistently exhibited higher values than Group 1. This effect size, while statistically significant, suggests a modest practical significance, consistent with findings by Tahir et al. [16], who noted similar effect sizes in their assessments of software measurement practices.

In assessing the validity of our study's findings, several potential threats must be considered. First, internal validity could be compromised by selection biases or non-random assignment of participants to groups, which might influence outcomes if the groups are not equivalent at baseline. This concern echoes the challenges identified by Calvo and Beltrán [18], who highlighted the importance of adaptability in risk assessment frameworks. Additionally, while our measures are standardized, they might not fully capture the constructs in question, leading to potential measurement errors, a limitation also noted in the work of Philippou et al. [19].

Externally, the generalizability of our results may be limited by the specific sample used, which may not accurately represent the broader population. This limitation is particularly relevant in light of the findings from Falco and Robiolo [20], who emphasized the importance of contextualizing metrics for practical implementation. Furthermore, the reliance on statistical assumptions, such as those inherent in t-tests and ANOVA, may not hold true across all datasets, potentially affecting the robustness of our conclusions.

Lastly, the effect size, though statistically significant, was small, suggesting that while differences between groups are present, their practical significance may be modest. This aligns with the observations made by Hsueh et al. [21], who discussed the need for careful interpretation of metrics in educational settings. Together, these analyses provide a comprehensive understanding of our findings in relation to existing literature, highlighting both the contributions and limitations of our study.

VIII. CONCLUSION

This paper has critically examined the impact of the GQM framework on SE exam outcomes. The analysis reveals that implementing the GQM framework significantly enhances the clarity and focus of the learning objectives, which in turn positively affects student performance in exams.

The lessons learned from this study suggest that the structured approach of GQM not only aids educators in designing more effective assessments but also helps students in aligning their study strategies to meet specific learning goals. The findings underscore the potential of the GQM framework as a powerful tool in educational settings, particularly in disciplines that require high levels of analytical and problem-solving skills like SE.

Furthermore, the statistical evidence supports the hypothesis that systematic goal setting within educational frameworks can lead to improved educational outcomes. This insight is crucial for educators seeking methods to enhance instructional quality and for educational institutions aiming to boost academic performance.

In light of these findings, future research should d address the following:

- Investigating the GQM framework's applicability across various disciplines beyond SE could provide insights into its versatility and effectiveness in different educational contexts.
- Longitudinal studies are needed to assess the sustained impact of the GQM framework on student learning outcomes and curriculum development over time.
- Examining the potential for integrating the GQM framework with educational technologies, such as learning management systems, may improve data collection and analysis for continuous improvement.

REFERENCES

[1] F. N. Colakoglu, A. Yazici, and A. Mishra, "Software product quality metrics: A systematic mapping study," IEEE Access, vol. 9, pp. 44647-44670, 2021.

[2] T. Galli, F. Chiclana, and F. Siewe, "Software product quality models, developments, trends, and evaluation," SN Computer Science, vol. 1, pp. 1-24, 2020.

[3] A. Yasin, R. Fatima, L. Wen, W. Afzal, M. Azhar, and R. Torkar, "On using grey literature and Google Scholar in systematic literature reviews in software engineering," IEEE Access, vol. 8, pp. 36226-36243, 2020.

[4] T. Tahir, T. Ghulam Rasool, and C. Gencel, "A systematic literature review on software measurement programs," Information and Software Technology, vol. 73, pp. 101-121, 2016.

[5] V. R. Caldiera, V. R. Basili, and H. D. Rombach, "The Goal Question Metric Approach," in Encyclopedia of Software Engineering, 1994, pp. 528-532.

[6] A. Janes and G. Succi, Lean Software Development in Action, Springer Berlin Heidelberg, 2014. doi: 10.1007/978-3-642-00503-9. [Online]. Available: https://doi.org/10.1007/978-3-642-00503-9.

[7] R. Van Solingen, V. Basili, G. Caldiera, and H. D. Rombach, "Goal question metric (GQM) approach," Encyclopedia of Software Engineering, 2002.

[8] A. I. Gonzalez-Tablas Ferreres, K. Wouters, B. Ramos Alvarez, and A. Ribagorda Garnacho, "EVAWEB: A web-based assessment system to learn X.509/PKIX-based digital signatures," IEEE Trans. Educ., vol. 50, no. 2, pp. 112–117, May 2007.

[9] E. Guzman and R. Conejo, "Self-assessment in a feasible, adaptive web-based testing system," IEEE Trans. Educ., vol. 48, no. 2, pp. 688–695, Nov. 2005.

[10] National Commission for Academic Accreditation and Assessment, "Handbook for Quality Assurance and Accreditation in Saudi Arabia: Part 2," Version 3, Oct. 2015. [Online]. Available: http://www.kfupm.edu.sa/deanships/dad/Documents/AAC/NCAAA%20Documents/H2.%20Handbook%20Part%202.pdf

[11] National Qualifications Framework for Higher Education in the Kingdom of Saudi Arabia, 2009. [Online]. Available: https://www.mu.edu.sa/sites/default/files/National%20Qualifications%20Framework%20for%20HE%20in%20KSA.pdf

[12] K. McLaughlin, S. Coderre, W. Woloschuk, and H. Mandin, "Does blueprint publication affect students' perception of validity of the evaluation process?" Adv. Health Sci. Educ., vol. 10, pp. 15–22, 2005.

[13] P. Bridge, J. Musial, R. Frank, T. Roe, and S. Sawilowsky, "Measurement practices: Methods for developing content-valid student examinations," Med. Teach., vol. 25, no. 4, pp. 414–421, 2003.

[14] Z. Meng, C. Zhang, B. Shen, and Y. Wei, "A GQM-based approach for software process patterns recommendation," in Proc. SEKE, 2017, pp. 370–375.

[15] A. M. Aldahmash and A. Gravell, "Measuring success in agile software development projects: a GQM approach," 2018.

[16] T. Tahir, G. Rasool, W. Mehmood, and C. Gencel, "An evaluation of software measurement processes in Pakistani software industry," IEEE Access, vol. 6, pp. 57868–57896, 2018.

[17] M. Shojaeshafiei, "Analytic hierarchy process-based fuzzy measurement to quantify vulnerabilities of web applications," International Journal of Computer Networks & Communications (IJCNC), vol. 12, 2020.

[18] M. Calvo and M. Beltrán, "Applying the Goal, Question, Metric method to derive tailored dynamic cyber risk metrics," Information & Computer Security, vol. 32, no. 2, pp. 133-158, 2024.

[19] E. Philippou, S. Frey, and A. Rashid, "Contextualising and aligning security metrics and business objectives: A GQM-based methodology," Computers & Security, vol. 88, p. 101634, 2020.

[20] M. Falco and G. Robiolo, "Product Quality Evaluation Method (PQEM): To understand the evolution of quality through the iterations of a software product," Int. J. Software Eng. Appl. (IJSEA), vol. 12, 2021.

[21] N.-L. Hsueh, J.-J. Wang, and D. Bilegjargal, "Building learning analysis system with GQM methodology and ELK stack," J. Internet Technol., vol. 24, no. 2, pp. 379–387, 2023.

[22] N. M. Minhas, T. R. Koppula, K. Petersen, and J. Börstler, "Using goal–question–metric to compare research and practice perspectives on regression testing," J. Softw. Evol. Process, vol. 35, no. 2, p. e2506, 2023.

AUTHOR'S PROFILE

Reem Alnanih is an associate professor of Computer Science in Faculty of Computing and Information Technology (FCIT), at KAU, Jeddah, Saudi Arabia. She received her B.Sc. and M.Sc. degrees from Computer Science Department. She got her Ph.D. in Computer Science from Concordia University, Montreal, Canada. She currently serves as the Vice Dean of Scientific Research for Female at KAU. Her research interests include software engineering, HCI, quality measurement and related evaluation techniques.

Critical Success Factors of Microservices Architecture Implementation in the Information System Project

Mochamad Gani Amri, Teguh Raharjo, Anita Nur Fitriani, Nurman Rasyid Panusunan Hutasuhut
Master of Information Technology-Faculty of Computer Science, University of Indonesia, Jakarta, Indonesia

***Abstract*—Microservice Architecture (MSA) promises enhancements in information systems, including improved performance, scalability, availability, and maintenance. However, challenges during the design, development, and operations phases can hinder successful deployment. This research presents a case study of one of the leading telecommunications companies in Indonesia, which encountered a three-month delay in implementing its microservices architecture (MSA). The study aims to provide actionable insights for the company to enhance its MSA deployment and contribute to academic knowledge by offering a structured approach to evaluating critical success factors (CSFs) in similar contexts. Through a literature review, twenty-one factors were identified and categorized into four groups: (1) Organization, (2) Process, (3) Systems & Tools, and (4) Knowledge, Skills & Behavior. The Analytical Hierarchy Process (AHP) was used to evaluate the priority of each factor based on survey data from project executors and software development practitioners. The findings indicate that the Organization category is the most crucial, with (1) Top Management Support, (2) Clear Vision, and (3) Adequate Resources being the top three CSFs for MSA implementation.**

***Keywords*—Microservice architecture; software architecture; critical success factors; analytical hierarchy process**

I. Introduction

Microservice Architecture (MSA) has emerged as a prevailing trend in software development in recent years [1]. The advent of microservice architecture gained traction in the early 2010s, serving as a countermeasure to the complexities encountered in monolithic application development, deployment, and scalability [2]. Software development teams and organizations across various sectors have embraced microservice architecture for the construction and management of their applications [3]. MSA's adoption spans multiple industries, from technology to e-commerce and banking, and across different development environments, from cloud computing to on-premise solutions [4].

MSA delineates a monolithic application into a collection of smaller, isolated, and interconnected services [5]. Each of these services bears distinct responsibilities and operates independently from others, thereby enabling development teams to work in isolation and reduce the overall system complexity [6]. MSA's popularity stems from its greater flexibility in development, the capability for independent scaling and updating of components, and the facilitation of system management and security [7].

Despite the improvements and benefits offered by MSA implementation, O'Reilly published research [8] that identified potential challenges (pains) arising at various implementation phases of MSA (design, development, and operations) such as overcoming existing mindset, decomposing functionality and integration to legacy system as the top three challenges. Cross-phase issues in design, development, and operation have been identified, including architecture, management, monitoring, testing, and others [9]. Several researchers have also explored the benefits (gains) and challenges (pains) of MSA deployment [9,10].

As a case study, one of Indonesia's leading telecommunications companies is improving its workforce management system to enhance availability, scalability, and maintainability. One of the initiatives being undertaken was migrating from monolithic to microservices-based architecture. However, during the development process, challenges arose, causing the project delivery to be delayed by three months beyond the initial target. Based on the interview conducted with one of the technical managers, the complexity of the current monolithic system and lack of microservice experience at the organizational level made the duration of important aspects such as development and testing take longer than expected. This case problem aligns with research and practical evidence, which suggests that MSA is not a panacea, as it presents several challenges during its implementation phases [11]. About 9% of respondents experienced complete failure (not successful at all), and 37% reported limited success (some success) in their MSA implementation efforts [8].

Previous studies have researched the challenges and benefits that can be referred to as a candidates for the factors that affect MSA implementation. First [6], offers an insightful historical perspective on the evolution of microservices, highlighting key transitions that have shaped modern software engineering practices. Second [10], provides valuable empirical evidence from industry on the practical challenges of adopting microservices, directly supporting my analysis of deployment differences between public cloud and on-premise systems. Third [11], delves into the current obstacles and future directions for microservices, enriching the discussion of ongoing challenges and emerging trends.

The significant benefits offered by MSA implementation, coupled with the challenges faced during its execution, form

the basis of this study. Previous researchers have laid a solid foundation of knowledge on the benefits, problems, and challenges of implementing MSA, yet there has been no structured collection of critical success factors (CSFs) within specific case studies, especially in Indonesia. Hence, this study aims to identify and address the research question:

RQ: What are the critical success factors of microservice architecture implementation in the information systems project?

The author also argues that the success of an initiative or program is intrinsically linked to an organization's internal capabilities. These capabilities are comprised of four components: (1) Organization, (2) Processes, (3) Systems & Tools, and (4) Knowledge, Skills & Behaviors [12]. These components form the basis for categorizing the critical factors identified in this study.

II. Research Methods

This study is organized into three stages, as illustrated in Fig. 1. Step 1 involves the identification of Critical Success Factors (CSFs) through a systematic literature review (SLR) and the categorization of these factors based on capability categories. Step 2 encompasses the development and execution of a survey designed to collect quantitative data, serving as the foundation for the prioritization process. Step 3 entails the evaluation of priority levels utilizing the Analytic Hierarchy Process (AHP), a methodology that facilitates multi-criteria decision-making.

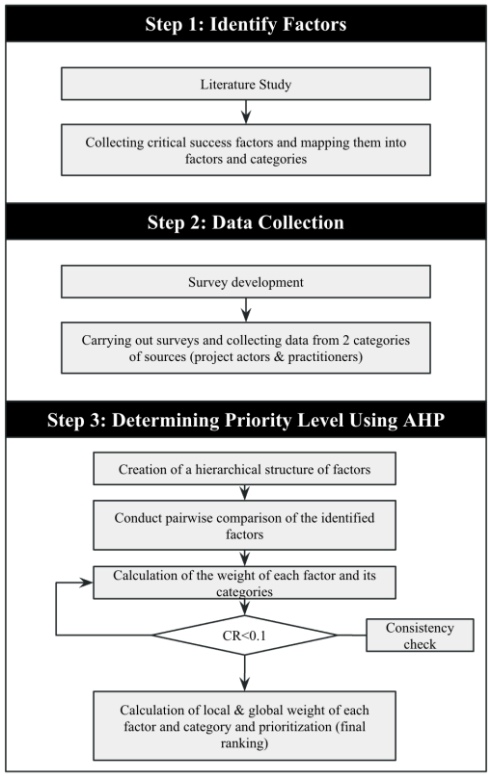


Fig. 1. Research method.

III. Identify Success Factors

A. Literature Study

In the initial stage of this study, the investigators conducted a literature study to identify the Critical Success Factors (CSFs) in the implementation of Microservice Architecture (MSA). Table I illustrates that the literature study commenced with a manual search through the electronic data source (EDS) Google Scholar, resulting in ten relevant studies.

TABLE I. Research Selection

Stages of Literature Study	Electronic Data Source (EDS)			
	Google Scholar (Manual)	Science Direct (SLR)	IEEE Xplore (SLR)	Total
String Execution	-	95	318	413
Study Extraction	-	20	48	68
Study Screening & Selection	10	11	8	29

For further research, two additional EDS, namely ScienceDirect and IEEE Xplore, were selected for systematic literature review (SLR) utilizing a three-step method.

- String Execution: The execution of search strings was performed according to the specifications outlined in Table II. These two data sources were targeted for the SLR and successfully gathered 413 research articles from journals and conferences.
- Study Extraction: The 413 research articles underwent a further filtering process based on titles & keywords to determine their relevance for in-depth reading. This resulted in 68 papers, with the primary aim being to avoid research focused on technical implementations in the form of application solutions that were outside the search for factors.
- Study Screening & Selection: A total of 29 studies, comprising 19 SLR studies (11 from ScienceDirect and eight from IEEE Xplore) plus ten manually identified studies, were advanced for closer examination. This involved quickly reading the abstracts, research findings, discussions, and conclusions to identify whether they contained fsactors relevant to MSA implementation.

TABLE II. SLR Research Specifications

Serarch Key Words	("microservices" OR "microservices architecture") AND ("factors" OR "success" OR "failure" OR "challenges")
Section	Title, Abstract & Author Keywords
Publication Type	Research Articles or Conferences
Publication Year	2019-2024

B. Define the Success Factors

In this section, the researchers elucidate the Critical Success Factors (CSFs) identified from the literature study. A total of 21 critical success factors were discovered for the

implementation of Microservice Architecture (MSA), described as follows.

1) Top management support: The support exhibited by top leaders of an organization towards a specific initiative or project encompasses recognition, resource allocation, and active involvement from top management in guiding, encouraging, and supporting the success of the initiative. Researchers concur that the success of information system implementation is inseparable from comprehensive support from top management within an agency or company [33-34].

2) Clear vision: Research [35] adds a clear vision as a critical success factor for information system implementation into its framework. A clear vision includes the goals to be achieved, available resources, timelines, systems to be implemented, vendors, and possible methods needed to achieve these goals.

3) Organizational culture: The implementation of MSA can drive organizational culture change through the division of teams into smaller units, reducing dependencies and communication between teams. This can also lower costs in the management process [10]. A survey conducted [30] mentions that MSA results in better development and deployment with scalability compared to monolithic architecture; however, there are challenges where organizational culture becomes a primary factor in MSA adoption. Organizational culture also significantly influences company performance and effectiveness, employee morale and productivity, and its ability to attract, motivate, and retain talented individuals [36].

4) Adequate resources: Project managers and software architects are the main focus of high development costs in MSA implementation due to developers lacking experience in MSA implementation [22]. The need to design a good MSA architecture to optimize the performance of MSA-based applications requires adequate software architecture resources [29]. In this research context, adequate resources include the technical, financial, and human resources allocated for the implementation process.

5) Project management: Effective project management is a critical need for every information system project initiative [31]. Project management encompasses the processes of planning, organizing, directing, and controlling the resources used to achieve the project objectives within specified time, budget, and scope limits. Requirements Engineering (RE) and non-RE tasks/sub-tasks, such as overseeing the transition of requirements to code, coordination, or project management, are regarded as essential skills in the era of MSA [43]. The synchronization between the teams is also a challenge in developing microservice architecture due to the complex dependencies and communication of services [47].

6) Change management: Managing resistance in the implementation process of a new initiative is an important factor; organizations must have good planning for projects to run effectively [31]. The process of managing change within an organization involves understanding, preparation, execution, and monitoring of changes to achieve the desired outcomes and minimize potential negative impacts.

7) Training and education: The main goal of training is to enhance the skills, knowledge, and competency levels of all users within the organization. Training should be provided and considered as part of the implementation process because it influences collective beliefs about the benefits of the implemented system [31]. Microservices practitioners should have competencies in at least one, ideally multiple, of the core MSA roles (web developer, DevOps engineer, or data engineer). The three main collections of competencies indicate the primary technical skills that practitioners who want to work with microservices should have [45].

8) Vendor selection: The process of selecting vendors that will collaborate with an organization to provide specific products, services, or solutions. This involves the evaluation and selection of vendors that best meet the needs and requirements of the organization. Careful selection of vendors, products, and services is necessary because failure to do so successfully can be very costly, as shown by several reported failure cases [25].

9) Agile methodology: The development of distributed microservice-based systems can adopt the agile approach as a system development method; therefore, communication is crucial for knowledge transfer among teams [24]. The software development process with agile methodology focuses on adaptive team collaboration, responsiveness to changes, and the delivery of high-value products. Organizations found agile development compatible with microservice-based architectures. The common view was that the bounded context of each microservice works well in agile development [47].

10)System modeling: The process of creating models that visually or conceptually represent a system. The goal is to understand, analyze, design, and construct systems in a systematic and structured manner, for example, using UML. Modeling diagrams are used to describe various aspects of MSA-based systems. Results show that Architecture and Functional Flow diagrams are commonly used to depict a high-level view of MSA [28]. A study found that the lack of a formal representation of domain models is a challenge for software modeling and emphasizes even simple and informal diagrams such as UML can be used to represent domain models as long as they help communicate design insights [42].

11)API management and standardization: API management involves managing the API lifecycle, including planning, development, testing, deployment, and maintenance of APIs (API Contract, API Versioning, Open API Standard, etc.). Further research worth conducting is on the pains felt due to exploiting APIs to enable microservice services to communicate [9]. The microservice API evolution process suffers from the loose coupling between services and leads to communication overheads and backward compatibility necessity [46].

12)Service communication: Communication between various microservices within a system (Discovery of services,

Replicated Service Instances, Load Balancing, Replicating data, Remote Calls, Relation Between Tables, REST, Event Driven). There are facts that intercommunication between API-based systems also poses real issues [9]. In a microservice-based system, to overcome the challenge of complex communication and dynamicity at runtime in failure detection, distributed tracing data is used [44].

13)Infrastructure automation: When infrastructure automation technology is applied, the deployment of microservices becomes as straightforward as monolithic systems [10]. Activities encompassing infrastructure automation processes include CI/CD, tools (version control, build, test, deploy), and testing tactics (unit, API, integration & contract testing).

14)Monitoring and logging: The application of tools for monitoring, information gathering, and data analysis to ensure optimal performance, detect issues, and support accurate decision-making. A major challenge of implementing microservices is the intrinsic complexity of monitoring applications composed of a large, dynamically evolving, and heterogeneous number of components [9]. There are also challenges to detecting failures in microservice-based systems due to the inherent characteristics of such systems, including complex communications, frequent updates, dynamicity at runtime, and complex log management [44]. Monitoring is of paramount importance to continuously drive development and testing, starting from the feedback collected from the field [48].

15)Cloud computing adoption: Data storage, processing, and resource management over the internet using infrastructure provided by cloud service providers such as IaaS, PaaS, and SaaS. SaaS Cloud-native based on microservices offers customers flexible infrastructure opportunities while leveraging the economies of scale provided by cloud providers and multi-tenancy architecture [27].

16)System architecture design: The expertise in planning the overall structure and components of a system. This involves identifying system needs, selecting appropriate technologies, modeling components, and arranging interactions between those components. In the process, considering microservices-based architecture design identifies parts (partitions) of applications in limited contexts is not easy; microservices yield benefits from implementing this bounded context [9].

17)Database architecture and design: Determining the appropriate database model, designing schema, defining tables and their attributes, and constructing relationships among them. Creating databases that can efficiently store and retrieve data, ensure data integrity, and meet system requirements. The design of a database per service pattern allows for independence among MSA services by equipping each service with its storage (if needed) [9].

18)System security: Knowledge related to practices for protecting computer systems, networks, or software applications from unauthorized access, data breaches & cyber threats. System security aims to ensure the confidentiality, integrity, and availability of system resources and data. Security also poses issues at the design time, especially due to access control (ACL) and the proliferation of endpoints [9] (Table III).

TABLE III. CRITICAL SUCCESS FACTORS BASED ON LITERATURE STUDIES

ID	Factors	Studies
F1	Top Management Support	L13 [32], L14 [33]
F2	Clear Vision	L15 [34]
F3	Organizational Culture	L9 [10], L11 [30] L16 [35]
F4	Adequate Resources	L2 [22], L10 [29]
F5	Project Management	L12 [31], L24[43], L28[47]
F6	Change Management	L12 [31]
F7	Training and Education	L12 [31], L26[45]
F8	Vendor Selection	L5 [25]
F9	Agile Methodology	L4 [24], L28[47]
F10	System Modeling	L8 [28], L23[42]
F11	API Management & Standardization	L1 [9], L9 [10], L11 [30], L17 [36], L27[46]
F12	Service Communication	L1 [9], L2 [22], L6 [26], L11 [30], L17 [36], L18 [39], L21 [40], L25[44]
F13	Infrastructure Automation	L1 [9], L2 [22], L6 [26], L8 [28], L9 [10], L11 [30], L18 [37], L20 [39], L21 [40]
F14	Monitoring & Logging	L1 [9], L2 [22], L6 [26], L8 [28], L9 [10], L11 [30], L18 [39], L21 [40], L25[44], L29[48]
F15	Cloud Computing Adoption	L7 [27], L10 [29]
F16	System Architecture Design	L1 [9], L2 [22], L8 [28], L11 [30], L17 [36], L18 [37], L21 [40]
F17	Database Architecture & Design	L1 [9], L9 [10], L11 [30], L17 [36], L18 [39], L21 [40]
F18	System Security	L1 [9], L6 [26], L8 [28], L11 [30], L17 [36], L21 [40], L22 [41]
F19	DevOps Culture	L8 [28], L9 [10], L11 [30], L19 [38]
F20	Microservices Experience	L2 [22], L2 [23], L10 [29], L11 [30], L18 [37]
F21	System & Software Design Patterns	L8 [28], L11 [30], L19 [38], L21 [40]

19)DevOps culture: The integration of software development (Dev) practices with IT operations (Ops) to create an efficient, sustainable workflow focused on the rapid delivery of business value. The combination of MSA and DevOps brings several other benefits, including increased software release cadence, system reliability and scalability, resilience in case of failures, and decentralized team

management to control application development [28]. In the technological aspect, containerization also enables DevOps to achieve faster deployment implementation than VMs as it is not efficient to run each microservice on a separate VM due to its long startup time and increased resource usage [41].

20)Microservices experience: Knowledge and experience in designing, developing, and managing application architectures using the microservices approach. Research [22] highlights higher development costs, possibly due to developers undertaking microservice development for the first time.

21)System and software design patterns: Several MSA design patterns have been identified. The most recurring design patterns during the MSA implementation process in DevOps are Circuit Breaker and migration patterns, followed by Observer, Load Balancer, Scalability, and Deployment [28].

According to Strategy, and a unit of PwC focused on business strategy consulting, capabilities are comprised of four components as illustrated in Fig. 2. Organization, Process, Tools & System, and Knowledge, Skills & Behaviors [12]. Each component is described as follows:

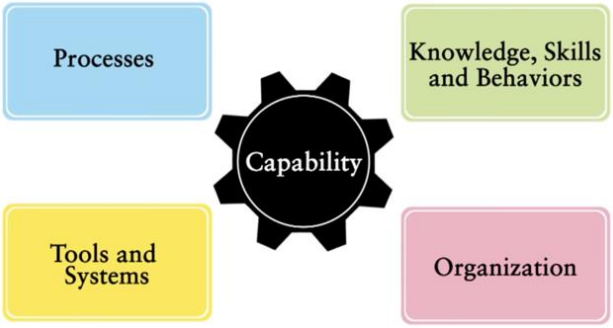


Fig. 2. Four parts of capability according to Strategy and PWC.

a) Organization: The entire structure and design of a company, including governance, management, and its operational model. Organizations with strong capabilities can identify, develop, and leverage competitive advantages in a competitive market. They can transform the knowledge, skills, and behaviors of individuals or groups into sustainable organizational capabilities that create value for all stakeholders involved.

b) Process: The steps or activities used by a company to create value for customers. In the context of an organization, processes aim to achieve efficiency, effectiveness, and desired outcomes. This is crucial to design, manage, and continuously improve processes to enable the organization to operate better, enhance customer satisfaction, and achieve competitive excellence.

c) Tools and system: The right combination of relevant systems and tools can strengthen the capabilities of an organization or individual, facilitating the coordination, efficiency, and monitoring needed to achieve desired goals. It is important to match systems and tools with the specific needs of the organization or individual and ensure they align with established strategies and objectives.

d) Knowledge, skills, and behaviors: In many situations, the success of an individual or group does not solely depend on the knowledge they possess but also on the skills applied in practice and behaviors that support good performance. These interrelated knowledge, skills, and behaviors often become a focus in the development of individuals and human resource management strategies to effectively achieve organizational objectives.

C. Decision Modeling Using AHP

The Analytic Hierarchy Process (AHP) is a widely recognized method for solving complex decision problems developed by Saaty [13]. It breaks down any complex issue into multiple sub-problems using AHP in terms of hierarchy levels, where each level represents a set of criteria or attributes relative to each sub-problem.

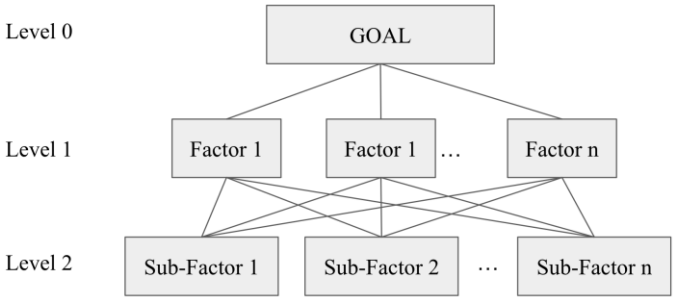


Fig. 3. AHP methodology structure as a framework for complex decision-making problems.

For instance, in an AHP model, Fig. 3 would illustrate the top level (level 0) of the hierarchy representing the goal of the problem, followed by the middle level (level 1) representing strategic and operational factors, and the last level (level 2) typically representing alternative actions or sub-factors that must be considered to achieve the goal. AHP organizes feelings, intuition, and logic in a structured approach to decision-making, which has proven beneficial in environments largely composed of intangible attributes.

AHP enables an individual to organize a system and its environment into interacting elements and then synthesize these by measuring and ranking the impact of these elements on the overall system [13,14]. AHP comprises four phases.

- Phase 1: Problem structuring and designing the hierarchy structure of factors and subfactors.
- Phase 2: Data collection based on pairwise comparisons through expert surveys. In the data collection process, interviews must be conducted with experts actively involved in problem resolution [16]. Subsequently, a pairwise comparison matrix A is formed as follows:

$$A = (a_{ij}), (i, j = 1 \ldots n) \quad (1)$$

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} = \begin{pmatrix} 1 & a_{12} & \dots & a_{1n} \\ 1/a_{12} & 1 & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ 1/a_{1n} & 1/a_{2n} & \dots & 1 \end{pmatrix} \quad (2)$$

- Phase 3: Determining the priority weights of each factor and performing a consistency check.

In the third phase, the focus is on identifying the relative importance values of the factors themselves within a specific category rather than comparing them to factors or categories elsewhere [17]. This can be achieved by calculating their local weights using (3) [16].

$$W_{Ai} = \frac{\left(\prod_{j=1}^{n} a_{ij}\right)^{1/n}}{\sum_{i=1}^{n}\left(\prod_{j=1}^{n} a_{ij}\right)^{\frac{1}{n}}} \quad (3)$$

Subsequently, the global weights, which indicate the relative importance of factors among all factors, are calculated as follows: the global weight of a CSF (Critical Success Factor) category is equal to its local weight. For CSFs, the global weight is the product of the local CSF weight and the local weight of the related category. This step concludes with a consistency test through the calculation of the Consistency Ratio (CR) as follows: calculation of the largest eigenvalue.

Calculate the biggest Eigen Value using (4).

$$A\omega = \lambda_{max}\omega \quad (4)$$

Calculate CR value using (5).

$$CR = \frac{CI}{RI} \quad (5)$$

- Phase 4: Analyzing the priority weights and determining solutions for the problem.

In the fourth phase, the priority weights calculated from the pairwise comparisons are analyzed to derive the best solutions for the problem. These weights represent the relative importance of each criterion or alternative, helping to prioritize options based on the decision-maker's preferences.

IV. Result and Discussion

A. Result

After identifying the critical success factors (CSFs) impacting the implementation of MSA, a hierarchical structure was constructed, and the upper categories of these factors were established. The categories are Organization, Process, Tools & System, and Knowledge, Skills & Behaviors, which are specifically explored in this study. The four categories and their respective factors are illustrated in Fig. 4.

Once the hierarchy/categorization was established, comparative assessment was employed to ascertain the relative significance of these factors. A pairwise comparison matrix was then developed to prioritize the comparative assessments into ratio scale measurements. In this study, a nine-point scale was applied to compare the significance of each factor pair, as detailed in Table IV. The evaluation consists of five values: 1, 3, 5, 7, and 9, representing equal importance, slight importance over another, essential or strong importance, demonstrated importance, and absolute/extreme importance, respectively. Intermediate assessments were expressed through the use of even numbers 2, 4, 6, and 8.

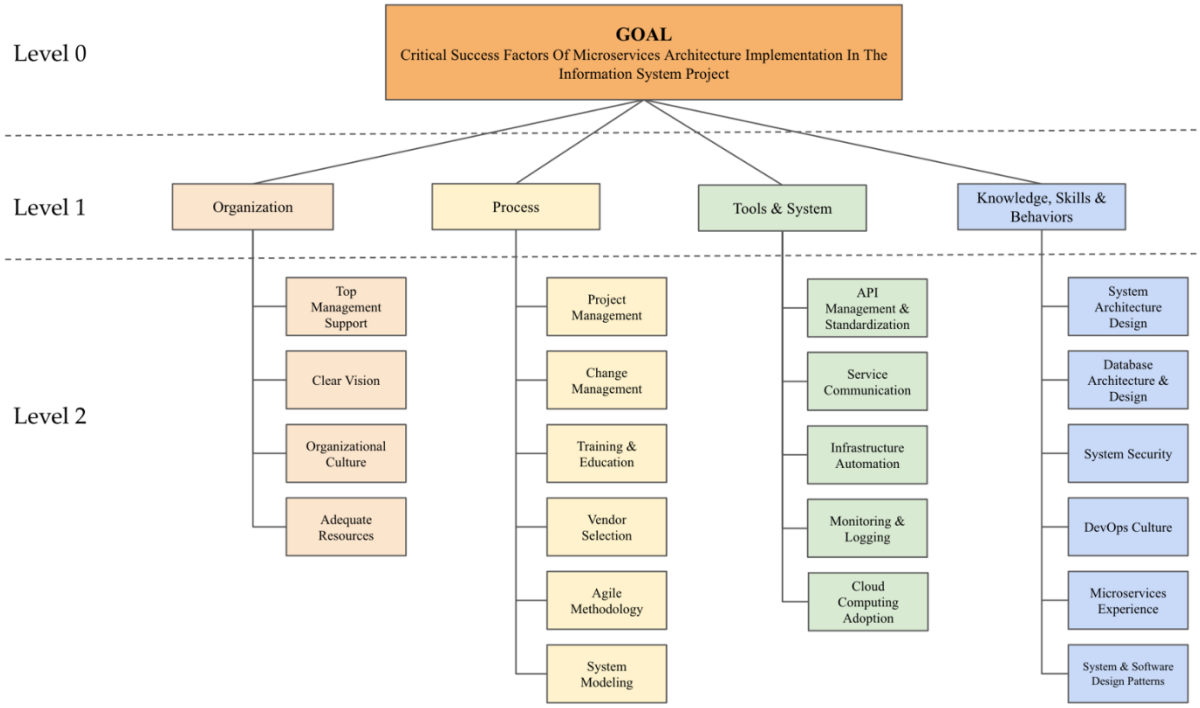


Fig. 4. Microservice architecture implementation success factors hierarchical structure.

Experts were asked to perform pairwise comparisons of factors. These comparisons were then used to create a pairwise comparison matrix. An example of such comparison is management support versus a clear vision, posing the question, "How significant is top management support compared to a clear vision?" If experts responded with "demonstrated importance," a value of 7 was assigned in the matrix. Conversely, if experts indicated that a clear vision holds greater importance than top management support, a value of 1/7 was added to the matrix. The involvement of experts in this prioritization was documented in a pairwise comparison survey.

TABLE IV. SCALE MEASUREMENT

i compare with j	a_{ij}	a_{ji}
Equally Importance (EI)	1	1/1
Moderately Importance (MI)	3	1/3
Strongly More Importance (SMI)	5	1/5
Very Strongly More Importance (VSMI)	7	1/6
Extremely More Importance (EMI)	9	1/7
Equally Importance (EI)	1	1/1

After completing the comparison matrix, researchers proceeded to calculate priority weights, consistency index, random consistency index, and the consistency ratio among the factors using the pairwise comparison matrix. Table V displays the random consistency index provided by [18]. It is crucial to note that when calculating the consistency ratio, a threshold must not be exceeded. To ensure desirable outcomes, it is advised that the consistency ratio is within the range of 0 and 0.1, especially for matrices exceeding a 4x4 dimension, as recommended by Saaty [18]. If the calculated consistency ratio is equal to or less than the acceptable value, this indicates that the comparative assessments represented in the matrix have a consistent level of consistency.

TABLE V. RANDOM CONSISTENCY INDEX

Size of Matrix	Random Consistency Index
1	0
2	0
3	0.58
4	0.90
5	1.12
6	1.24
7	1.32

In this study, the process of calculating priorities was divided into four according to the number of categories, and the results from the consistency index check were as follows: Organization (Cat 1) 0.0931, Process (Cat 2) 0.0924, System & Tools (Cat 3) 0.0972, and Knowledge, Skills & Behaviors (Cat 4) 0.0822. These results indicate that the overall process of calculating priorities is consistent.

Table VI displays the results of the priority weights and rankings for each category and the factors associated with them. The Category column contains weight values representing the importance level of each category. The four categories derived from the supporting parts of the capabilities required by any company or institution, in the context of this study, refer to capabilities in implementing MSA. Analysis revealed that the Organization category is most crucial, with a weight of 0.4844 in the implementation of MSA. The Knowledge, Skills & Behaviors category ranks second with a weight of 0.2240, followed closely by the Process and Tools & System categories with weights of 0.1615 and 0.1302, respectively.

In the Factor column, there are two weight values: local weight as a priority within the category scope of the factor and global weight as the basis for the Priority column as the factor's ranking compared to all studied factors. In the Organization category, the Top Management Support factor is deemed most important, with a local weight of 0.5011. For the Process category, the Project Management factor is the main priority with a local weight of 0.3046. In the Tools & System category, the highest local weight of 0.3938 is for the API Management & Standardization factor, and in the last category, Knowledge, Skills & Behaviors, the Microservices Experience factor is considered most important with a local weight of 0.2300. Globally, the top three critical success factors are Top Management Support, Clear Vision, and Adequate Resources.

B. Discussion

This study focuses on the process of identifying, defining, and evaluating critical success factors (CSFs) for the implementation of Microservices Architecture (MSA) at one of the leading telecommunications companies in Indonesia. Each factor will be mapped into categories based on an institution's or company's capabilities to implement MSA. This mapping process is crucial as a high-level overview for the management at the company to understand which internal capabilities need to be enhanced for similar initiatives to be carried out effectively and efficiently in the future. A literature study was conducted by combining a manual search of studies deemed relevant to the topic discussed, supplemented with the application of a systematic literature review (SLR) to enrich the determinants of successful MSA implementation.

Table VII displays the demographics of the experts who assisted in the prioritization process of the critical success factors (CSFs) in this study. We were supported by four experts, consisting of two types: those who were involved and those who were not directly involved with the project implementation. The presence of these two types of experts is expected to achieve a balance that supports better objectivity in the prioritization process conducted through surveys. The experts were selected with several selection criteria such as more than five years of experience in the software field, having experience in developing systems based on MSA, and being certified in software architecture or IT service management.

TABLE VI. SUMMARY OF WEIGHTING AND PRIORITY LEVEL OF FACTORS

Categories			Factors					
ID	*Name*	*Weight*	*ID*	*Name*	*Local Weight*	*Local Rank*	*Global Rank*	*Priority*
Cat 1	Organization	0.4844	F1.1	Top Management Support	0.5011	1	0.2427	1
			F1.2	Clear Vision	0.2630	2	0.1274	2
			F1.3	Organizational Culture	0.0768	4	0.0372	9
			F1.4	Adequate Resources	0.1591	3	0.0771	3
Cat 2	Process	0.1615	F2.1	Project Management	0.3046	1	0.0492	6
			F2.2	Change Management	0.1700	3	0.0274	14
			F2.3	Training & Education	0.0830	6	0.0134	20
			F2.4	Vendor Selection	0.1557	4	0.0251	16
			F2.5	Agile Methodology	0.1072	5	0.0173	18
			F2.6	System Modeling	0.1795	2	0.0290	13
Cat 3	Tools & System	0.1302	F3.1	API Management & Standardization	0.3938	1	0.0513	5
			F3.2	Service Communication	0.2853	2	0.0371	10
			F3.3	Infrastructure Automation	0.1346	3	0.0175	17
			F3.4	Monitoring & Logging	0.1039	4	0.0135	19
			F3.5	Cloud Computing Adoption	0.0824	5	0.0107	21
Cat 4	Knowledge, Skills & Behaviors	0.2240	F4.1	System Architecture Design	0.1550	4	0.0347	11
			F4.2	Database Architecture & Design	0.1312	5	0.0294	12
			F4.3	System Security	0.2004	2	0.0449	7
			F4.4	DevOps Culture	0.1675	3	0.0375	8
			F4.5	Microservices Experience	0.2300	1	0.0515	4
			F4.6	System & Software Design Patterns	0.1160	6	0.0260	15

TABLE VII. EXPERTS DEMOGRAPHICS

Criteria		Total	Percentage
Age	20-30	1	25%
	30-50	3	75%
	> 50	0	0%
Education	Bachelor	2	50%
	Master	2	50%
	Doctoral	0	0%
Experience	2-5 years	1	25%
	5-10 years	3	50%
	> 10 years	1	25%
Job Title	Technical Staff	3	75%
	Manager	1	25%
	Executive	0	0%
Project Involvement	Yes	2	50%
	No	2	50%

The Analytic Hierarchy Process (AHP) was employed to derive local and global weights as benchmarks for determining the priority levels of each category and critical success factors (CSFs) for the implementation of Microservices Architecture (MSA). This study produced a taxonomy of critical success factors (CSFs), taking into account both global and local weights. The identified success factors were categorized into four distinct sections; every section offers several insights for enhancing the MSA implementation process at the company.

Fig. 5 presents the outcome of the global weighting process for all determining factors of success using AHP as the method for complex decision-making. It can be concluded that the factor Top Management Support significantly outperforms other factors. Following in second and third order are the factors of Clear Vision and Adequate Resources respectively. An interesting finding is that these three factors fall within the same category of Organization. This justifies that the Organization category, displayed in Fig. 6, also ranks first from a capability perspective and significantly surpasses other categories, such as Knowledge, Skills & Behaviors in second place. The results of this study address the research question.

For future research, we acknowledge several limitations of this study, particularly in determining the critical success factors (CSFs), although the literature review utilized the Systematic Literature Review (SLR) method. The results from these factors have not undergone empirical testing for more objective factor selection. Approaches such as the frequency approach, found in [19,20], could be employed. Furthermore, the implementation of an expert judgment committee to evaluate the factors could be conducted to ensure that each factor is clearly defined [21].

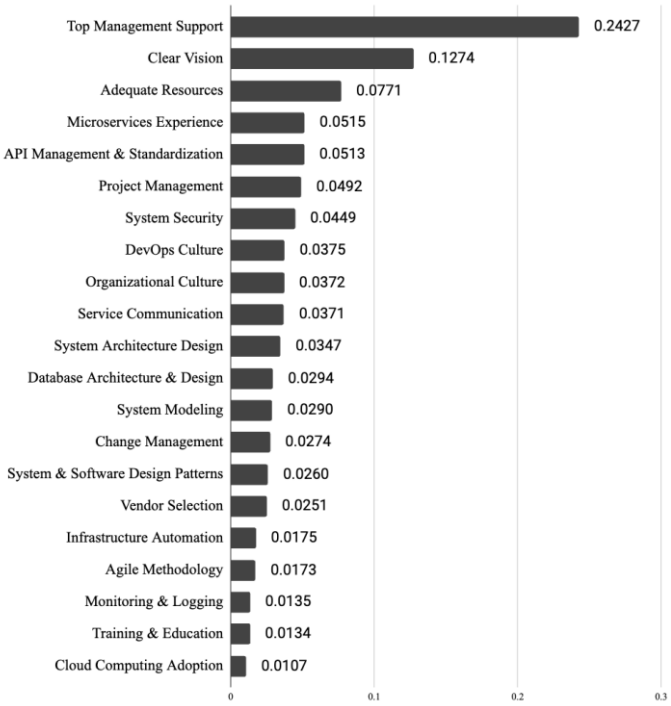


Fig. 5. List of overall implementation success factors ordered by global weight.

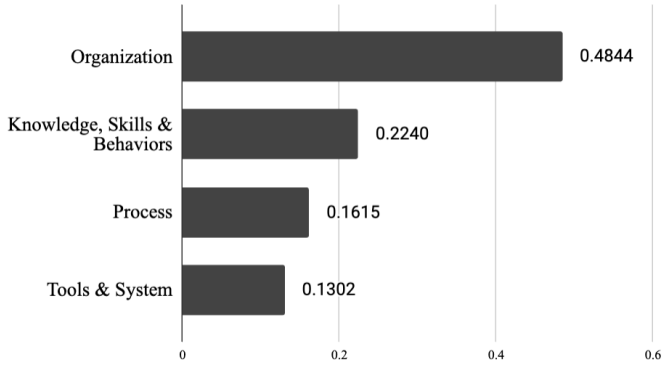


Fig. 6. List of categories based on capabilities ordered by weight.

V. Conclusion and Future Work

In this study, 21 critical success factors (CSFs) were successfully identified from the literature review process and categorized into four categories: (1) Organization, (2) Process, (3) Tools & System, and (4) Knowledge, Skills & Behaviors. To perform the evaluation process of the priority level of each factor and category, the Analytical Hierarchy Process (AHP) was selected as the decision-making methodology based on survey data conducted among project participants and practitioners.

The conclusions from the results of this study, as depicted in Table VI, indicate that from a categorical perspective, Organization ranks first as the most important. The top three sub-factors of critical success for the implementation of Microservices Architecture (MSA) are: (1) Top Management Support, (2) Clear Vision, and (3) Adequate Resources. When examining the priority of each category, based on local weights, it is evident that in the Organization category, the primary priority factor is Top Management Support (F1.1). For the second category, Process, the main priority is factor Project Management (F2.1). In the category Tools & System (F3.1), the factor API Management & Standardization is the priority, and for the last category, Knowledge, Skills & Behaviors, Microservices Experience (F4.5) ranks as the top factor.

To ensure the successful implementation of Microservices Architecture (MSA) in the organization, it is crucial to enhance leadership engagement, strategic clarity, and resource allocation. Strengthening top management support is vital, as their active involvement and oversight are essential for maintaining momentum and addressing challenges effectively. Additionally, articulating a clear vision and developing a robust communication strategy will ensure that teams are aligned and working towards common objectives, thereby minimizing misalignment and inefficiencies.

Furthermore, the company should prioritize the allocation of sufficient resources, both financial and human, to support the MSA initiative. This includes investing in the recruitment and upskilling of personnel to develop the necessary expertise in microservices, as well as ensuring access to essential tools and systems. By addressing these critical areas, the company can leverage the identified critical success factors to mitigate potential risks, optimize the execution of the project, and fully realize the benefits of MSA, including enhanced scalability, flexibility, and operational efficiency.

For future work, it is recommended that further research be conducted to explore the long-term impacts of the identified critical success factors (CSFs) on the sustainability and scalability of Microservices Architecture (MSA) implementations. Additionally, conducting comparative studies across different industries and organizational sizes would be valuable to determine whether the CSFs identified in this study are universally applicable or if they vary based on context.

References

[1] Smith, D., Jones, M., Williams, R., & Johnson, L. (2020). Embracing Microservices: A Comprehensive Overview of Microservice Architecture and Its Impact on Software Development. Journal of Software Engineering and Applications, 13(4), 123-139.

[2] Gupta, S. (2016). Microservices. Procedia Computer Science, 78, 278-283.

[3] Johnson, R., & Ho, A. (2019). Adopting microservices at Netflix: Lessons for architectural design. IEEE Software, 36(1), 39-45.

[4] Lewis, J., & Fowler, M. (2014). Microservices. online: https://martinfowler.com/articles/microservices.html (Accessed 20 May 2023)

[5] Newman, S. (2015). Building Microservices: Designing Fine-Grained Systems. O'Reilly Media, Inc, 551.

[6] Dragoni, N., Giallorenzo, S., Lafuente, A. L., Mazzara, M., Montesi, F., Mustafin, R., & Safina, L. (2017). Microservices: yesterday, today, and tomorrow. Present and Ulterior Software Engineering, 195-216.

[7] Balalaie, A., Heydarnoori, A., & Jamshidi, P. (2016). Microservices architecture enables devops: Migration to a cloud-native architecture. IEEE Software, 33(3), 42-52.

[8] Loukides, Mike. & SwoyerLewis, Steve. (2020). Microservices Adoption in 2020. online: https://www.oreilly.com/radar/microservices-adoption-in-2020/ (Accessed 20 May 2023)

[9] Jacopo Soldani, Damian Andrew Tamburri, Willem-Jan Van Den Heuvel. (2018). The pains and gains of microservices: A Systematic

gray literature review. The Journal of Systems and Software 146 (2018) 215-232.

[10] Xin Zhou, Shanshan Li a, Lingli Cao, He Zhang, Zijia Jia, Chenxing Zhong, Zhihao Shan, Muhammad Ali Babar. (2023). Revisiting the practices and pains of microservice architecture in reality: An industrial inquiry. The Journal of Systems & Software 195 (2023) 111521

[11] Jamshidi, P., Pahl, C., Mendonca, N.C., Lewis, J., Tilkov, S., 2018. Microservices: the journey so far and challenges ahead. IEEE Softw. 35 (3), 24-35. doi:10.1109/MS.2018.2141039.

[12] Strategy& (2014). What is a capability?. online: https://www.strategyand.pwc.com/gx/en/about/media/videos/2015-and-older/what-is-a-capability.html (Accessed 20 May 2023)

[13] Saaty TL. What is the analytic hierarchy process? In:Mathematical Models for Decision Support. Berlin, Heidelberg: Springer; 1988:109-121.

[14] Saaty TL. The analytic hierarchy process. New York: McGraw Hill; 1980.

[15] Saaty TL. Decision making for leaders. Pittsburgh: RWS Publications; 1990.

[16] Cheng, E.W. and Li, H. (2001),"Analytic hierarchy process: an approach to determine measures for business performance",Measuring Business Excellence, Vol. 5 No. 3, pp. 30-37.

[17] Shi, H., Peng, S.Z., Liu, Y. and Zhong, P. (2008),"Barriers to the implementation of cleaner production in chinese SMEs: government, industry and expert stakeholders perspectives",Journal of CleanerProduction, Vol. 16 No. 7, pp. 842-852.

[18] Saaty, T.L., 1986. Axiomatic foundation of the analytic hierarchy process. Manag. Sci. 32 (7), 841-855.

[19] Khan, A. A., & Shameem, M. (2020). Multicriteria decision-making taxonomy for DevOps challenging factors using analytical hierarchy process. Journal of Software: Evolution and Process, 32(10). https://doi.org/10.1002/smr.2263

[20] Akbar, M. A., Naveed, W., Mahmood, S., Alsanad, A. A., Alsanad, A., Gumaei, A., & Mateen, A. (2020). Prioritization Based Taxonomy of DevOps Challenges Using Fuzzy AHP Analysis. IEEE Access, 8, 202487-202507. https://doi.org/10.1109/ACCESS.2020.3035880

[21] Sambasivan, M., & Fei, N. Y. (2008). Evaluation of critical success factors of implementation of ISO 14001 using analytic hierarchy process (AHP): a case study from Malaysia. Journal of Cleaner Production, 16(13), 1424-1433. https://doi.org/10.1016/j.jclepro.2007.08.003

[22] Lenarduzzi, V., Lomio, F., Saarim√ § ki, N., & Taibi, D. (2020). Does migrating a monolithic system to microservices decrease the technical debt? In Journal of Systems and Software (Vol. 169). Elsevier Inc. https://doi.org/10.1016/j.jss.2020.110710

[23] Auer, F., Lenarduzzi, V., Felderer, M., & Taibi, D. (2021). From monolithic systems to Microservices: An assessment framework. Information and Software Technology, 137. https://doi.org/10.1016/j.infsof.2021.106600

[24] Taibi, D., Lenarduzzi, V., Pahl, C., & Janes, A. (2017). Microservices in agile software development: A workshop-based study into issues, advantages, and disadvantages. ACM International Conference Proceeding Series, Part F129907. https://doi.org/10.1145/3120459.3120483

[25] Stefanou, C. J. (n.d.). Association for Information Systems Association for Information Systems AIS Electronic Library (AISeL) AIS Electronic Library (AISeL) The Selection Process of Enterprise Resource Planning (ERP) The Selection Process of Enterprise Resource Planning (ERP) Systems Systems The Selection Process of Enterprise Resource Planning (ERP) Systems. https://aisel.aisnet.org/amcis2000/418

[26] Karabey Aksakalli, I., √áelik, T., Can, A. B., & TekiÃánerdo*f*üan, B. (2021). Deployment and communication patterns in microservice architectures: A systematic literature review. Journal of Systems and Software, 180. https://doi.org/10.1016/j.jss.2021.111014

[27] Nordli, E. T., Haugeland, S. G., Nguyen, P. H., Song, H., & Chauvel, F. (2023). Migrating monoliths to cloud-native microservices for customizable SaaS. Information and Software Technology, 160. https://doi.org/10.1016/j.infsof.2023.107230

[28] Waseem, M., Liang, P., & Shahin, M. (2020). A Systematic Mapping Study on Microservices Architecture in DevOps. Journal of Systems and Software, 170. https://doi.org/10.1016/j.jss.2020.110798

[29] Hasan, M. H., Osman, M. H., Admodisastro, N. I., & Muhammad, M. S. (2023). Legacy systems to cloud migration: A review from the architectural perspective. Journal of Systems and Software, 202. https://doi.org/10.1016/j.jss.2023.111702

[30] Waseem, M., Liang, P., Shahin, M., di Salle, A., & M√° rquez, G. (2021). Design, monitoring, and testing of microservices systems: The practitioners' perspective. Journal of Systems and Software, 182. https://doi.org/10.1016/j.jss.2021.111061

[31] Merhi, M. I. (2021). Evaluating the critical success factors of data intelligence implementation in the public sector using analytical hierarchy process. Technological Forecasting and Social Change, 173. https://doi.org/10.1016/j.techfore.2021.121180

[32] Merhi, M. I., & Leighton, J. (2015). A process model leading to successful implementation of electronic health record systems. In Int. J. Electronic Healthcare (Vol. 8).

[33] Young, R., & Jordan, E. (2008). Top management support: Mantra or necessity? International Journal of Project Management, 26(7), 713-725. https://doi.org/10.1016/j.ijproman.2008.06.001

[34] Gaardboe, R., Nyvang, T., & Sandalgaard, N. (2017). Business Intelligence Success applied to Healthcare Information Systems. Procedia Computer Science, 121, 483-490. https://doi.org/10.1016/j.procs.2017.11.065

[35] Warrick, D. D. (2017). What leaders need to know about organizational culture. Business Horizons, 60(3), 395-404. https://doi.org/10.1016/j.bushor.2017.01.011

[36] Waseem, M., Liang, P., Ahmad, A., Shahin, M., Khan, A. A., & M√° rquez, G. (2022). Decision models for selecting patterns and strategies in microservices systems and their evaluation by practitioners. 135-144. https://doi.org/10.1145/3510457.3513079

[37] Poniszewska-Maranda, A., MacIoch, J., Borowska, B., & Maranda, W. (2021). Mechanisms for Transition from Monolithic to Distributed Architecture in Software Development Process. Proceedings - IEEE Computer Society's Annual International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems, MASCOTS. https://doi.org/10.1109/MASCOTS53633.2021.9614287

[38] Nino-Martinez, V. M., Octavio Ocharan-Hernandez, J., Limon, X., & Perez-Arriaga, J. C. (2021). Microservices Deployment: A Systematic Mapping Study. Proceedings - 2021 9th International Conference in Software Engineering Research and Innovation, CONISOFT 2021, 24-33. https://doi.org/10.1109/CONISOFT52520.2021.00016

[39] Waseem, M., Liang, P., Marquez, G., & Salle, A. di. (2020). Testing microservices architecture-based applications: A systematic mapping study. Proceedings - Asia-Pacific Software Engineering Conference, APSEC, 2020-December, 119-128. https://doi.org/10.1109/APSEC51365.2020.00020

[40] Premarathna, D., & Pathirana, A. (2021). Theoretical framework to address the challenges in Microservice Architecture. Proceedings - International Research Conference on Smart Computing and Systems Engineering, SCSE 2021, 195-202. https://doi.org/10.1109/SCSE53661.2021.9568346

[41] Sultan, S., Ahmad, I., & Dimitriou, T. (2019). Container security: Issues, challenges, and the road ahead. IEEE Access, 7, 52976-52996. https://doi.org/10.1109/ACCESS.2019.2911732

[42] Zhong, C., Li, S., Huang, H., Liu, X., Chen, Z., Zhang, Y., & Zhang, H. (2024). Domain-Driven Design for Microservices: An Evidence-Based Investigation. IEEE Transactions on Software Engineering, 50(6), 1425–1449. https://doi.org/10.1109/TSE.2024.3385835

[43] Ayas, H. M., Hebig, R., & Leitner, P. (2024). The Roles, Responsibilities, and Skills of Engineers in the Era of Microservices-Based Architectures. 2024 IEEE/ACM 17th International Conference on Cooperative and Human Aspects of Software Engineering (CHASE), 13–23.

[44] Purfallah Mazraemolla, Z., & Rasoolzadegan, A. (2024). An effective failure detection method for microservice-based systems using distributed tracing data. Engineering Applications of Artificial

Intelligence, 133(PF), 108558. https://doi.org/10.1016/j.engappai.2024.108558

[45] Michael Ayas, H., Hebig, R., & Leitner, P. (2024). An empirical investigation on the competences and roles of practitioners in Microservices-based Architectures. Journal of Systems and Software, 213(March), 112055. https://doi.org/10.1016/j.jss.2024.112055

[46] Lercher, A., Glock, J., Macho, C., & Pinzger, M. (2024). Microservice API Evolution in Practice: A Study on Strategies and Challenges. Journal of Systems and Software, 215(May), 112110. https://doi.org/10.1016/j.jss.2024.112110

[47] Ünlü, H., Kennouche, D. E., Soylu, G. K., & Demirörs, O. (2024). Microservice-based projects in agile world: A structured interview. Information and Software Technology, 165(September 2023). https://doi.org/10.1016/j.infsof.2023.107334

[48] Giamattei, L., Guerriero, A., Pietrantuono, R., Russo, S., Malavolta, I., Islam, T., Dînga, M., Koziolek, A., Singh, S., Armbruster, M., Gutierrez-Martinez, J. M., Caro-Alvaro, S., Rodriguez, D., Weber, S., Henss, J., Vogelin, E. F., & Panojo, F. S. (2024). Monitoring tools for DevOps and microservices: A systematic grey literature review. Journal of Systems and Software, 208(November 2023), 111906. https://doi.org/10.1016/j.jss.2023.111906

Application Citespace Visualization Tool to Online Public Opinion Group Label: Generation, Dissemination and Trends

Jingyi Ju*
School of Media, Nanyang Institute of Technology, Nanyang 473000, Henan, China

***Abstract*—With the popularization of mobile Internet technology, the social and cultural environment provides a favorable communication environment for online news dissemination, leading to a highly ubiquitous phenomenon of group labeling communication events. This study explores the generation, dissemination, and evolution trends of group labeling in the online public opinion environment. This study crawled 20975 initial literature data included in the core database of Web of Science, obtained 9834 valid literature after several rounds of screening, and utilized CiteSpace 6.3 software to do metric data analysis and word frequency analysis on the above valid literature data and successively adopted the Analysis means of literature being co-cited, author co-citation, journal co-citation, keyword co-citation, and clustering to analyze group labeling. We have successively used the Analysis of literature co-citation, author co-citation, journal co-citation, keyword co-citation, clustering, etc., and disassembled the group labeling into the generation environment, dissemination process, and trend evolution. The study utilizes the disciplinary perspective of journalism and communication and social media platforms to assist it. Thus, it summarizes and reveals the interaction of group labels in cyberspace and the natural world and seeks to grasp the communication mechanism of the process of insight into this emerging discursive power.**

***Keywords*—Group labeling; online public opinion; labeled communication; communication mechanisms; visualization**

I. Introduction

The proliferation of mobile Internet technology has significantly reshaped the social and cultural landscape, providing a fertile ground for the pervasive phenomenon of online news dissemination and group labeling communication events [1]. This study is set against the backdrop of a globally connected digital society, where the number of Internet users has surged to 5.16 billion, as reported by the "Global Digital Report 2023". In China, this figure has reached 1.092 billion, with an Internet penetration rate of 77.5% [2]. The transition from traditional one-to-one communication models to the current one-to-many socialized network models has been accelerated by the iterative updates of Internet technology. This evolution has not only democratized and personalized information dissemination but also given rise to the intriguing phenomenon of group labeling, which is the focal point of this research [3].

While the existing literature has made strides in understanding the mechanics of online communication and the spread of information, there is a notable gap in the comprehensive analysis of group labeling within the online public opinion sphere [4]. Prior studies have often focused on isolated aspects of this phenomenon, neglecting the interconnectedness and dynamic nature of group labels as they evolve and disseminate across various social platforms [5]. This research aims to address these shortcomings by providing a holistic examination of the generation, dissemination, and trend evolution of group labeling. Through the application of CiteSpace 6.3 software and a variety of analytical methods, we seek to uncover the underlying communication mechanisms and the interplay between group labels in cyberspace and the natural world.

The structure of this paper is meticulously designed to guide the reader through a comprehensive exploration of the topic. Section II. Data collection and research methodology: We detail our research methodology, explaining the data collection and analysis strategies using CiteSpace 6.3. Section III. Analysis process and findings: we present our analysis and findings, discussing the results of bibliometric and keyword analyses. Section IV. Dissemination and trends in group labeling: We examine the broader impact of group labeling, discussing its spread across social platforms and its societal effects. Finally, we conclude by summarizing our findings and their implications for online communication in Section V. We also assess our study's contributions and suggest future research directions, reflecting on our insights and looking forward to further exploration.

II. Data Collection and Research Methodology

In the subsequent section, we delve into a thorough review of the existing literature to contextualize our research within the broader academic discourse. This comprehensive overview serves to elucidate the current state of knowledge, underscore the methodologies employed by preceding studies, and highlight the contributions of pivotal research to the field. We critically assess the merits and limitations of existing approaches, shedding light on the shortcomings that our study endeavors to overcome. By identifying the gaps in the literature, we position our research as a response to the need for a more nuanced understanding and a more sophisticated approach to analyzing online public opinion group labeling. Our review not only pays homage to the foundational works that have paved the way for our study but also sets the stage for the innovative aspects of our methodology, which we believe will significantly advance the field by addressing the identified shortcomings and expanding the current horizons of research.

A. Data Acquisition

Based on the complexity of Internet information data collection and the different focuses of Internet group labels in other periods, this study mainly adopts the literature analysis method to analyze the data. To ensure the professionalism, leadership, accuracy, and completeness of the data, this paper selects the global academic literature that has been included in the core database of Web of Science as the data source and searches for the subject terms or contents, including "Group tags" or" Dissemination Mechanisms," "Social Communications," "Hashtag communication." The subject area is journalism and communication, and the time is set as the natural year between 2013 and 2023. After several rounds of searching, the initial literature data of 20975 items are obtained. After the big data technology-assisted data collection, manual review is also indispensable; through the reading of literature topics and literature abstracts, after eliminating a series of meetings, notices, reports, and other relevant literature data unsuitable for inclusion in the study, selecting the platform open access resources, and ultimately retaining 9834 pieces of valid literature data after manual screening and organizing.

B. Research Methodology

Based on the effective data sources obtained after the above multiple screenings, this study, with the help of the co-occurrence, clustering and social network analysis system in the CiteSpace 6.3 software tool, constructs the visualization matrices of bibliometrics, textual co-occurrence, word-frequency clustering, and social network relational mapping required for this study, and identifies the group labeling, a research object, through the generalization and Analysis of the extracted particular data related research hotspots, and then form a comprehensive, multidimensional and interconnected knowledge system to show the connections and interactions between different disciplines and the field of journalism and communication [3].

III. Analysis Process and Findings

In this section, we underscore the distinctive merits of our proposed methodology by juxtaposing it with existing approaches within similar domains. Our method leverages advanced data analytics and machine learning algorithms, which offer superior precision and scalability compared to traditional models. This advantage is particularly evident in handling large datasets commonly encountered in the analysis of online public opinion. Furthermore, our approach incorporates a novel feature engineering technique that enhances the model's ability to capture subtle nuances in group labeling dynamics, thereby providing more insightful predictions and interpretations. When compared with state-of-the-art models in the field, our method demonstrates a significant improvement in both accuracy and computational efficiency, as evidenced by the rigorous experimental evaluation detailed in the subsequent section. This comparative analysis not only highlights the innovative aspects of our approach but also solidifies its potential to contribute to the broader discourse on online group labeling and public opinion analysis.

A. Analysis of Literature Releases

Literature analysis identifies and responds to the knowledge and theoretical frameworks within a research discipline or field. In this study, the data were collected from the WoS core database. The literature was searched and filtered by "Group tags" and "Social Communications." The number of valid documents retrieved was 9834 by December 31, 2023. The number of valid documents retrieved was 9834 by December 31, 2023. As of December 31, 2023, the number of valid documents retrieved is 9834 articles. When analyzing the literature, this paper analyzes the number of publications in-depth and the co-citation relationship of the literature.

1) Number of literature releases: In the academic world, the number of literature releases can directly reflect the research activity, research strength, and research results of research scholars within the field of study, as well as the development of academic research. Therefore, the higher the number of literature releases represents, the more active and fruitful academic research in the field. This study mainly counts the literature released in the natural years of 2013-2023, in Fig.1.

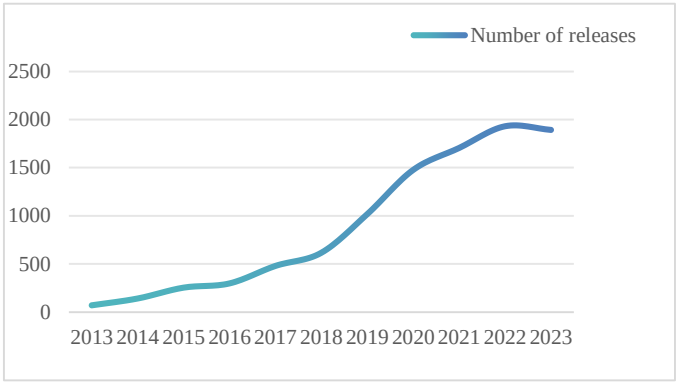


Fig. 1. Statistics on the number of literature releases, 2013-2023.

Based on the statistics and quantitative distribution of the number of literature releases, the literature releases within the field show a significant growth trend, starting from 69 in FY2013 and reaching 1,894 in FY2023. In particular, the growth rate of the number of annual publications has accelerated significantly from 2019 onwards, from 1,018 articles in FY2019 to 1,932 articles in FY2022, reflecting a significant increase in research activity and academic output within the field in recent years, which may be related to factors such as the development of Internet technology, new media forms, and digital technology. In addition, the number of articles issued in FY2022 and FY2023 accounted for 19.580% and 19.195% of the total, respectively, which indicates that these two years are the peak periods of academic research in this research field, which may be related to factors such as the increase in research funding, or policy support.

Upon further analyzing these data, it can be observed that the overall growth trend has remained the same between 2017 and 2019, despite fluctuations in the number of annual publications. To some extent, this growth phenomenon reflects the continuously growing research interest of research scholars in this field of study and the fresh exploration of its depth and breadth.

2) Literature co-citation analysis: Literature co-citation analysis is an important research method that reveals the structure of knowledge in a discipline. When two documents appear simultaneously in the reference list of the same cited document, they constitute a co-citation relationship. This means of Analysis can reveal the intrinsic connection between the literature, reflecting the knowledge base and developmental lineage within the field of study. Identifying and analyzing these co-cited documents makes it possible to locate documents central to the field and profoundly impact subsequent research. The cross-citation of literature demonstrates the dissemination and flow of knowledge over time and reflects the continuity of knowledge accumulation and the trajectory of disciplinary development [4]. If a paper is frequently cited within multiple research areas, it means that it has fundamental contributions to the theory and practice of the field. Co-citation analysis helps to identify this critical literature, thus providing a macro perspective for understanding the evolution of the discipline and current research trends. For the rapidly developing field of online public opinion, co-citation analysis is critical, as it can help research scholars quickly grasp the knowledge structure and research hotspots of the field and lay the foundation for further in-depth research.

In this study, an in-depth co-citation cluster analysis of scientific literature was conducted using CiteSpace software to deconstruct the knowledge architecture of the research area and gain insights into its developmental trends. The Analysis generated a network graph containing 1010 nodes and 3917 connectors, where nodes symbolize frequently co-cited literature and connectors reveal their interconnections. The network has a density value of 0.0077, reflecting that in such an extensive network, despite a large number of nodes, the connections between them are relatively sparse, a common characteristic of large network graphs (see Fig. 2).

Fig. 2. Literature co-citation visualization network mapping.

By observing the above-visualized network mapping, the network modularity Q-value reaches 0.7227, which, according to the modularity metric value standard, is not only much higher than the threshold of 0.3 in this study but also extremely close to the value of 1, which shows that there is a significant stable structure in the literature co-citation network [5]. In the literature co-citation map, it can be observed that several closely related subfields and a number of less closely related research topics are organized and connected around a particular core of knowledge within the research field. In addition, the Mean Silhouette value is as high as 0.8921, which is close to the perfect consistency 1. This is sufficient to show that the network knowledge structure of this research area is highly similar after the literature co-citation is clustered, further verifying the reliability and validity of the literature co-citation clustering.

B. Analysis of Study Authors

1) Author co-occurrence mapping: In Citespace software, we can explore the structure of author collaboration networks through specific parameter settings. In this study, we set the time slice to 1 year and chose "author" as the network node type to construct an author network structure graph. The graph comprises 520 nodes interconnected by 282 lines to form a panoramic knowledge graph [6]. However, the network density of this atlas is only 0.0021, showing that the overall network is very loose and lacks tightly cooperative clusters. In this network graph, nodes and connectivity are the two essential elements that make up the graph. Nodes represent individual authors, and their size is directly proportional to the number of articles published by an author, i.e., the larger the node, the richer the research output of that author. The connecting lines represent the cooperation relationship between authors, and their thickness represents the closeness of cooperation - the thicker the connecting lines are, the more frequent the collaboration between authors is in Fig. 3.

The author's collaboration network mapping generated by Citespace allows for a more intuitive identification of collaboration patterns and potential collaboration opportunities. Although the overall author collaboration network appears loose and does not form a large-scale close collaboration network, the localized collaboration activity still provides rich academic communication and collaboration opportunities [7]. Upon closer inspection, it is easy to find a certain degree of collaboration among authors, especially on a small scale, dominated by small teams of two to three people working together. This pattern of small-scale cooperation is a common phenomenon in academia, reflecting the close collaboration between individuals in academic research.

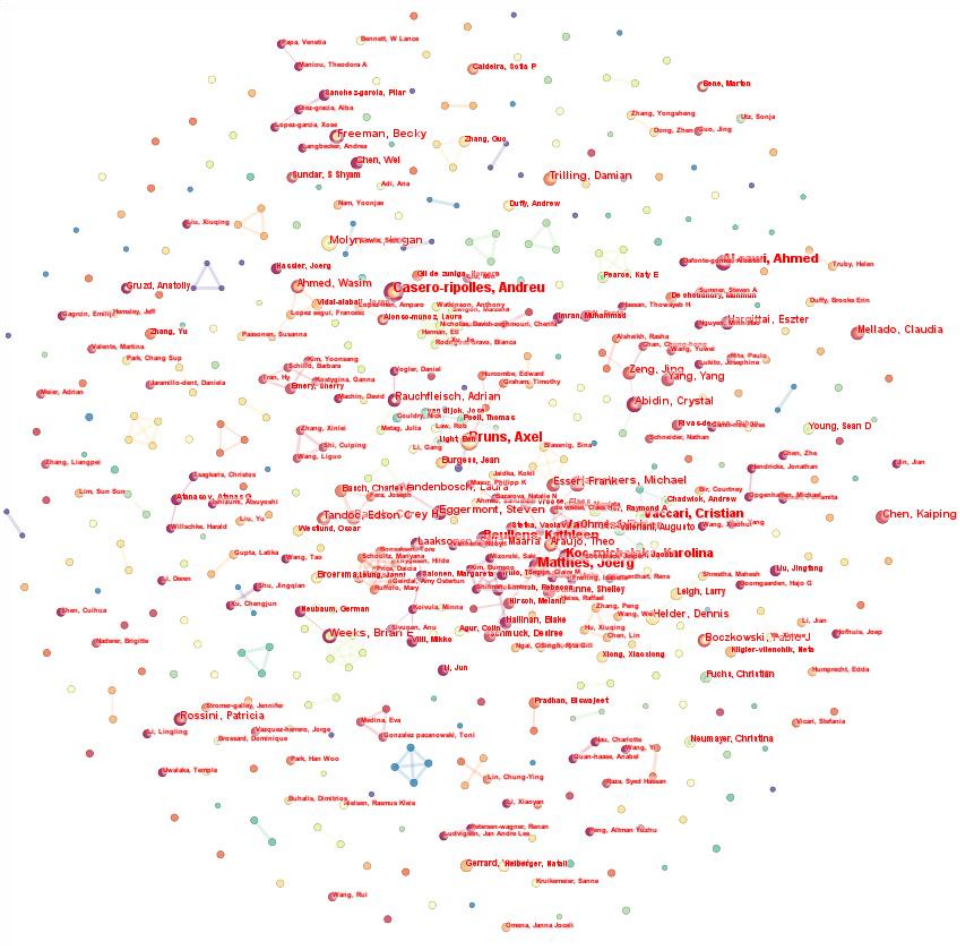


Fig. 3. Collaborating authors visualization network mapping.

This study conducted exhaustive statistics based on first-author literature publications. The data showed 520 authors published research papers in this field during 2013-2023. Among these authors, 76 published only one paper, accounting for 14.62% of the total number of authors. Although low, this percentage has demonstrated the research field's openness and inclusiveness to new scholars. Two hundred sixteen authors have published two papers, accounting for 41.54% of the total number of authors, and the more significant high percentage share reveals that research scholars in this field have a wide range of participation.

Combined with the time and frequency of literature publication, these authors may have a basic grasp and understanding of the research topic, and their research results may be more focused on the construction of theoretical frameworks or the exploration of specific issues, which, to a certain extent reflects the exploration and accumulation of knowledge in the field [8]. In addition, 141 authors published three papers, accounting for 27.12%, and this group of scholars provides continuous and stable research results in the field. Eighty-seven authors, or 16.73%, published more than four papers, indicating that this group of research scholars has in-depth research and significant academic contributions.

2) Analysis of core authors: Calculated according to Price's formula $M = 0.749 \times Nmax1/ 2$, in this study, the core authors $N = 16$ were brought into the formula calculation, which resulted in M equal to approximately 5.99, which means that the core authors should have at least six or more publications [9]. Thirty-two authors in our study had more than this standard of publications, Table I below.

Matthes-Joerg, Casero-ripples-Andreu, Bruns-Axel, Vaccari-Cristian, Van least-Peter, Koc-Michalska-Karolina, Al-rawi-Ahmed, Beullens-Kathleen and other authors have more than ten publications per capita. It shows their significant academic contribution and activity in the field. By analyzing the negativity of the authors and the structure of the collaborative network, these scholars occupy a central position in the overall research network and significantly impact academic discussions and knowledge dissemination.

Through the Analysis, it is easy to see that a stable research community has been formed in this research area centered on these 32 core authors. These authors' collaboration patterns are primarily focused on the same institutions or under the guidance of their supervisors, indicating the formation of a relatively close academic community. However, there is little cross-regional collaboration, which implies room for further development in promoting broader academic exchanges and cooperation.

C. Analysis of Journal Releases

1) Journal co-citation analysis: Journals are an essential platform for knowledge production in a research field or discipline, and the analysis and visualization of Journal co-citation is a way to reveal the centralized publication platform of literature and core journals [10]. Analyzing the co-citation frequency of core journals can effectively demonstrate the quality level of the journal articles. By choosing the node type as "cited journals" and selecting the journals with a frequency higher than 200, we obtained the co-citation network graph of journals consisting of 1368 nodes and 7228 connecting lines, with a Q-value of 0.7258, which indicates that the network clustering effect is good, and a Silhouette value of 0.9083, which suggests that the clustering result has a high credibility and reasonable results. Results have high credibility, and the results are reasonable in Fig. 4.

TABLE I. STATISTICS ON THE NUMBER OF LITERATURE RELEASES BY CORE AUTHORS

No.	Name and surname	Number of literature releases	No.	Name and surname	Number of literature releases
1	Matthes-Joerg	16	17	Mellado-Claudia	7
2	Casero-ripolles-Andreu	16	18	Rossini-Patricia	7
3	Bruns-Axel	15	19	Rauchfleisch-Adrian	7
4	Vaccari-Cristian	11	20	Zeng-Jing	7
5	Van Aalst-Peter	11	21	Chen-Kaiping	6
6	Koc-michalska-Karolina	11	22	Trilling-Damian	6
7	Al-Rawi-Ahmed	10	23	Boczkowski-Pablo J	6
8	Beullens-Kathleen	10	24	Vandenbosch-Laura	6
9	Molyneux-Logan	9	25	Ohme-Jakob	6
10	Abidin-Crystal	9	26	Hargittai-Eszter	6
11	Hameleers-Michael	9	27	Ahmed-Wasim	6
12	Basch-Corey H	8	28	Yang-Yang	6
13	Eggermont-Steven	8	29	Araujo-Theo	6
14	Weeks-Brian E	8	30	Esser-Frank	6
15	Freeman-Becky	8	31	Laaksonen-Salla-Maaria	6
16	Helder-Dennis	7	32	Tandoc-Edson C	6

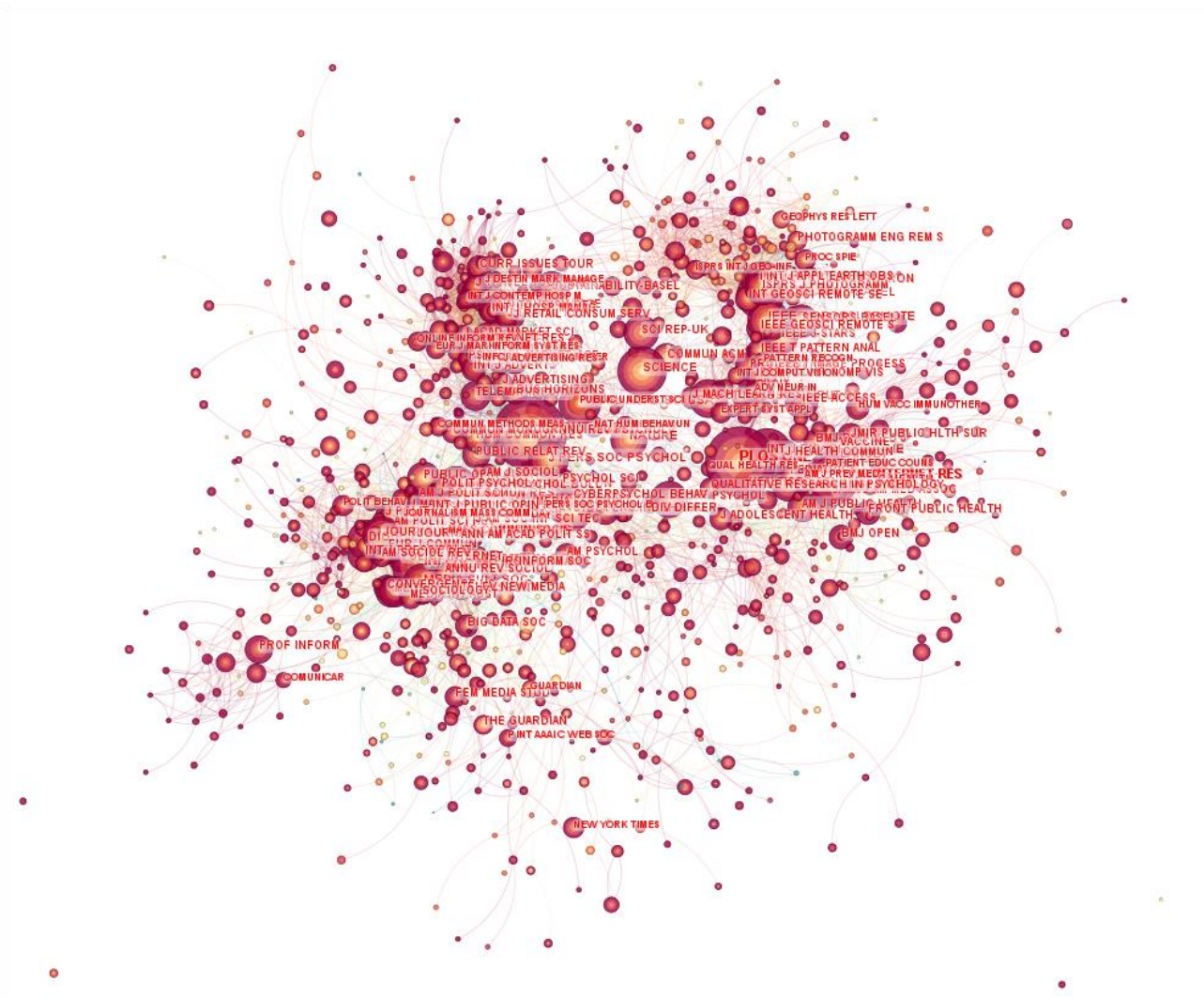


Fig. 4. Journal co-citation visualization network mapping.

In the visual mapping analysis of co-citations of academic journals, journals with significant impact in this field of study can be identified [11].CiteSpace6.3 uses specific visual coding to demonstrate the importance of journals, for example, through the color of different shades of the circle and the size of the circle as an indicator of the magnitude of the role of journals in the overall citation network. According to the journal co-citation visualization maps generated by CiteSpace 6.3 software, the maps of NEW MEDIA SOC, PLOS ONE, COMPUT HUM BEHAV, INFORM COMMUN SOC, J COMMUN, J COMPUT-MEDIAT COMM, SOC MEDIA SOC, J MED INTERNET RES, INT J COMMUN-US, and the prestigious SCIENCE journals ranked among the top in terms of citation frequency, with 2509, 1998, 1973, 1867, 1611, 1421, 1266, 1135 and 1041 citations, respectively. The number of citations more than 1,000 times not only reflects the high citation rate of these journals in this field of academic research in the academic community but also reacts to the core position of the above journals in the exchange of research fields in the academic community.

In addition, in this network mapping, NEW MEDIA SOC and PLOS ONE have the most extensive circle sizes and the most circle levels, indicating that they have the highest mediated centrality, which visually highlights that these two journals have a significant influence in research in group labeling. This highly mediated centrality also implies that these journals are essential platforms for publishing research results, providing rich academic research resources. They are also hubs of communication in academic dialog and knowledge dissemination, contributing to accumulating and developing knowledge in the field.

2) Analysis of core journals: The co-citation information of the journals searched out by CiteSpace 6.3 software was used for the econometric Analysis of the core journals in the field of this study by using the formula r_0=2 ln (eE*Y) calculated in conjunction with Bradford's law [12]. In this study, the value *Y*=2509 was estimated to be 27 journals in the core area, which accounted for 20% of the total number of journals, and they contained 29,655 articles, which accounted for 21.41% of the total literature in Table II.

TABLE II. CITATION FREQUENCY STATISTICS OF CORE JOURNALS

No.	Journal name	Citation frequency	No.	Journal name	Citation frequency
1	NEW MEDIA SOC	2509	15	IEEE T GEOSCI REMOTE	822
2	PLOS ONE	1998	16	POLIT COMMUN	819
3	COMPUT HUM BEHAV	1973	17	DIGIT JOURNAL	797
4	INFORM COMMUN SOC	1867	18	MEDIA CULT SOC	792
5	J COMMUN	1611	19	LECT NOTES COMPUT SC	789
6	J COMPUT-MEDIAT COMM	1421	20	J BROADCAST ELECTRON	780
7	SOC MEDIA SOC	1416	21	REMOTE SENS ENVIRON	755
8	J MED INTERNET RES	1266	22	Journalism Stud	735
9	INT J COMMUN-US	1135	23	JOURNALISM	719
10	SCIENCE	1041	24	J PERS SOC PSYCHOL	690
11	REMOTE SENS-BASEL	994	25	LANCET	686
12	COMMUN RES	955	26	NATURE	684
13	int j env res pub he	893	27	AM BEHAV SCI	666
14	P NATL ACAD SCI USA	842	28	J MASS COMMUN Q	649

In terms of the major journals in the core area, NEW MEDIA SOC carried the most significant number of articles, totaling 2509 articles, accounting for 1.81% of the total number of articles, followed by PLOS ONE (1998 articles), and COMPUT HUM BEHAV (1973 articles), with a share of 1.44% and 1.42%, respectively.

D. Keyword Analysis

In academic research, two core methods are usually used to identify the evolution of research hotspots and trends. The first is keyword co-occurrence analysis, an approach based on the original keywords recorded within the database, which serve as direct labels for the research literature, and by counting and analyzing the frequency of occurrence of these keywords, it is more intuitively responsive to the research hotspots and the main research directions in academia [13]. The second way, keyword cluster analysis, reveals the intrinsic connection between research topics by systematically grouping the associated keywords into clusters [14]. Next, this paper adopts these two keyword analysis methods to conduct an in-depth and comprehensive analysis of the "group labeling" research topic.

1) Keyword co-occurrence analysis: Keywords play a crucial role in academic papers; not only are they the concentrated manifestation of academic research positions and ideas, but high-frequency keywords also often reflect the research focus of the academic community in a certain period. To accurately capture these research hotspots, this paper adopts CiteSpace 6.3 software to conduct multiple rounds of screening on the selected sample data and then extract the relevant keywords. Subsequently, the extracted keywords were manually counted and sorted by word frequency through an Excel document to obtain a list of keywords reflecting the research trends in this research area in Fig. 5.

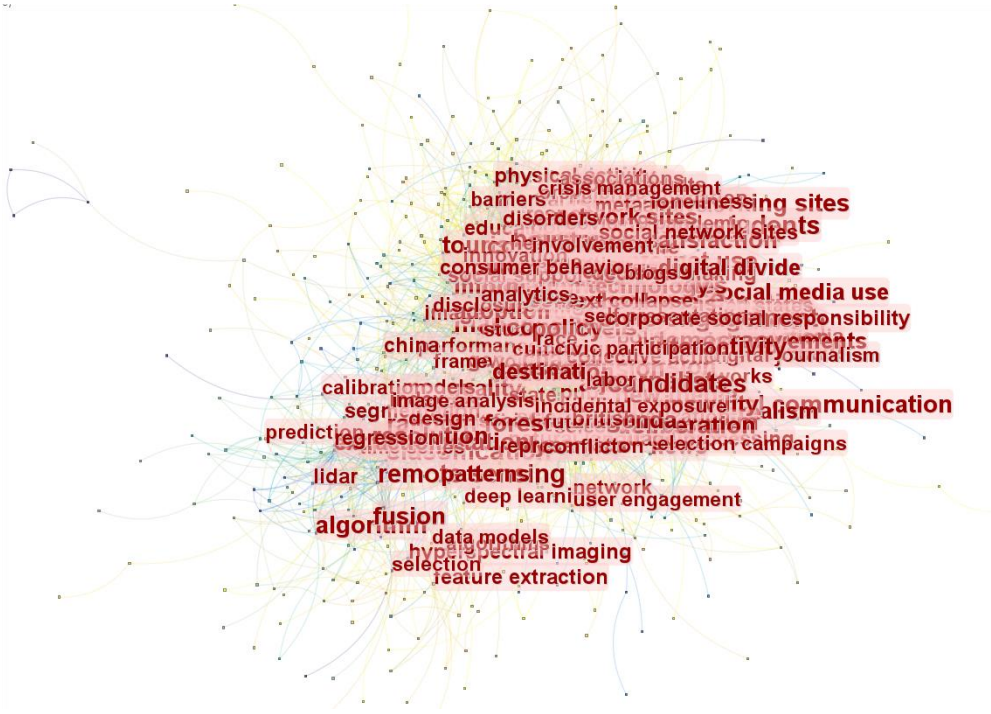


Fig. 5. Keyword co-occurrence visualization network mapping.

This research paper's core theme is identifying and analyzing group labels under online public opinion. CiteSpace 6.3 software can construct a keyword co-occurrence visualization network atlas with 763 nodes and 2854 connecting lines. Through visual observation, the node colors in the atlas show changes from dark green gradient to light green, red-light orange-light yellow, revealing the prosperous co-occurrence relationship between different keywords. The shades of these colors represent the importance of the nodes in the network; the darker the color, the more significant the keyword's centrality in the research domain. In addition, the size of the node also provides essential information for the study, which is proportional to the mediated centrality value of the keyword [15]. Keywords with higher mediational centrality values also have larger node sizes, indicating that the keyword links other keywords and forms the knowledge network's structure with solid control and influence. It facilitates the connection between other keywords and plays the role of the gas pedal in the evolution of academic research in Table III.

TABLE III. FREQUENCY STATISTICS OF KEYWORDS

No.	Byword	Frequency	No.	Byword	Frequency
1	social media	4202	30	technology	159
2	communication	668	31	machine learning	156
3	impact	536	32	sentiment analysis	154
4	information	526	33	public health	154
5	media	497	34	mental health	149
6	Twitter	496	35	trust	147
7	Facebook	459	36	networks	145
8	online	413	37	word of mouth	142
9	news	408	38	identity	139
10	Internet	389	39	consumption	133
11	model	344	40	quality	132
12	health	269	41	journalism	130
13	behavior	235	42	gender	125
14	perceptions	230	43	support	123
15	engagement	223	44	strategy	123
16	participation	217	45	coverage	118
17	classification	216	46	science	118
18	political communication	213	47	patterns	117
19	fake news	207	48	risk	116
20	social networks	207	49	algorithm	113
21	exposure	200	50	discourse	112
22	politics	181	51	Health communication	108
23	deep learning	178	52	adolescents	108
24	big data	175	53	performance	105
25	attitudes	171	54	satisfaction	105
26	content analysis	164	55	community	103
27	remote sensing	161	56	age	101
28	knowledge	160	57	care	100
29	management	159			

This paper sets the keyword list with a word frequency threshold more significant than 100. Thus, 57 primary keywords are screened out among thousands of keywords, and the cumulative frequency of these keywords reaches 52.57%, which accounts for more than half of the keyword co-occurrence visualization network mapping and meets the baseline of the Analysis [16]. Meanwhile, these high-frequency keywords help academic research scholars capture the main hotspots and trends in this research field and provide directions and focuses for further research.

When analyzing the research in this field in depth, the keyword co-occurrence mapping reflects some of the research hotspots. Social media, communication, impact, information, and media are the five most frequent keywords that form the cornerstone of the research in this field. In 2013, the term Internet appeared alongside social media, and both keywords continue to occur frequently in research that intersects with different disciplines in the field through 2024. The frequent co-occurrence of the keywords not only highlights the centrality of social media in the dissemination and communication of information but also demonstrates the widespread interest of research scholars in the impact of social media.

"Twitter," "Facebook," "online," "news," and "internet," as slightly less frequent keywords, represent media forms in the discipline of journalism and communication that are closely related to the development of Internet technology. The co-occurrence of these keywords reflects an in-depth academic study of how emerging social media platforms influence information dissemination and user behavior. Especially for globalized social media platforms such as Twitter and Facebook, researchers and scholars have focused on analyzing the psychological and behavioral effects of their information dissemination on audiences and users and then evaluating social media's communication patterns and effects.

The emergence of keywords such as "model," "health," "behavior," "perceptions," and "engagement" reveals the evolution of Internet user behavior and the formation of group dynamics in cyberspace. Internet users have shifted from single-individual behavior to group behavior, which has profoundly impacted society's functioning. Users have changed

from passive information receivers to active information producers and identify themselves and the groups they belong to through "group labeling." This change in identity has had a significant impact on the structure of information production, promoting the formation of a new decentralized communication structure and providing conditions for the democratization of group labels.

2) Keyword clustering analysis: The keywords were analyzed using CiteSpace 6.3 software for clustering combinations, and the keywords were re-categorized and identified by applying algorithms to the clusters. The initial clustering obtained 20 different categories, and after rigorous screening to eliminate categories with insufficiently tight or ineffective clustering, nine tags with significant clustering effects were finally retained by manual screening. These hashtags include but are not limited to 00#Social media, 01#Online review, 02#Populist communication, 03#Covid-19 pandemic, 04#Political participation, 05#Political participation, 06#Political participation, 07#Political participation. Political participation, 05#Social capital, 06#Covid-19 vaccination, 07#Driving factor, 08#Social media influencer, 09#Populist communication, 09#Covid-19 pandemic, 04#Political participation, 05#Social capital, 06#Covid-19 vaccination, 07#Driving factor, 08#Social media influencer, 08#Social media influencer (social media influencer) in Fig. 6.

After in-depth Analysis and re-aggregation of the keyword clustering results, three dominant research directions in this field were identified: social networks, user behavior and cognition, and communication patterns. Within the category of social networks, "social media" and "social capital" are particularly prominent. Social media is a core component of social networks and a critical force for social development, playing an increasingly significant role in the evolution of society. The popularity and penetration of social media are gradually changing users' daily lives and behavior patterns. In this era, technological advances have empowered users with more rights and voice, building a new type of social capital. The accumulation of this social capital not only reshapes the user's identity and behavioral perception but also has a far-reaching impact on social structure and personal interaction. As a socialization tool, the influence and effectiveness of social media have been increasing in modern society, and it has become one of the research hotspots that research scholars have paid attention to.

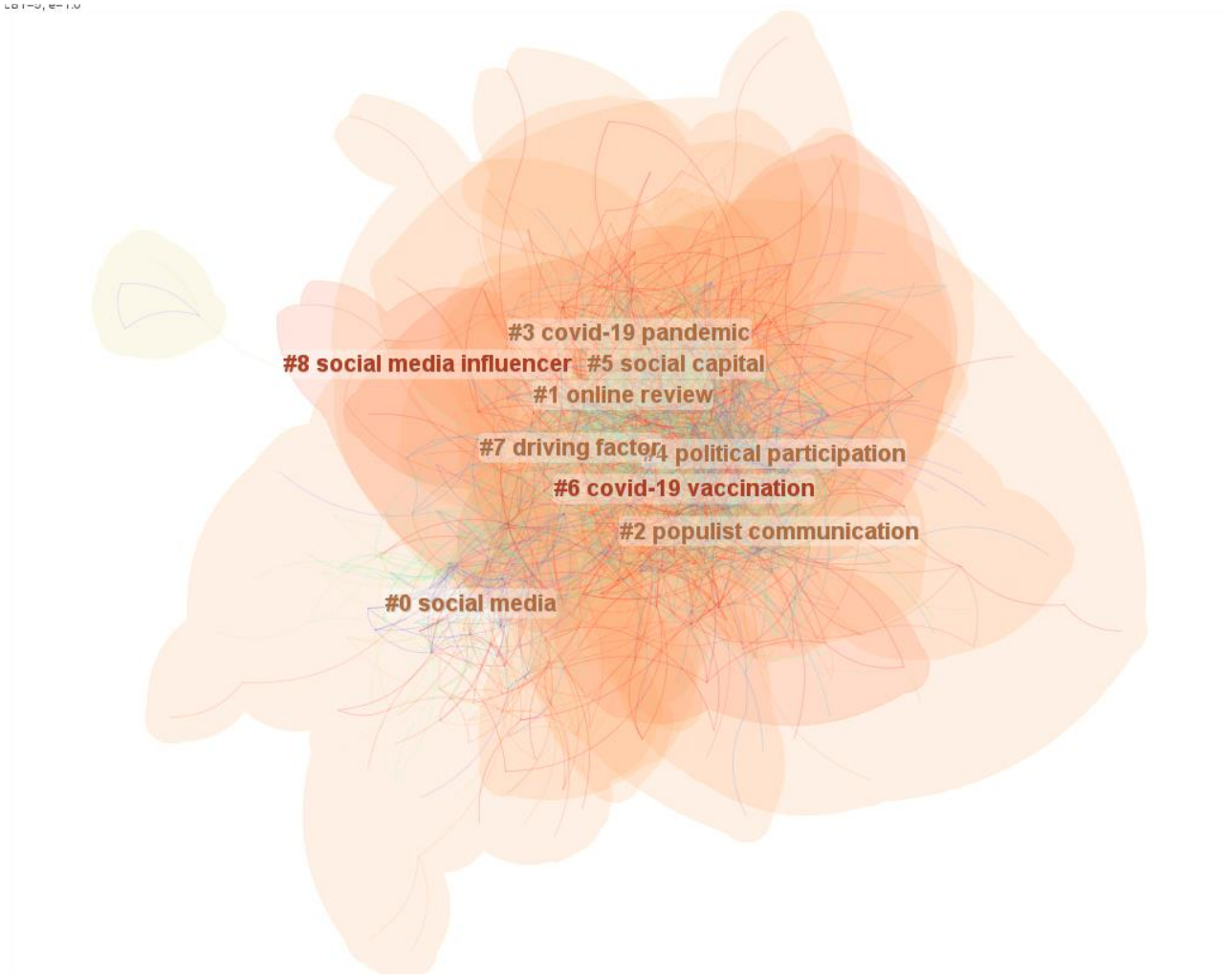


Fig. 6. Keyword clustering combination visualization network mapping.

The second category focuses on user behaviors and perceptions, covering the two cluster combinations of "Online review" and "Political participation." The evolution of technology has blurred the boundaries between online and offline spaces, reshaping users' daily habits and behavioral patterns. In this process, Internet users have formed groups with new "group labels" based on their active behaviors and new identities, and their group behaviors have broadly impacted social development. It is worth noting that there are two particular clusters in this category - "Covid-19 vaccination" (Covid-19 vaccination) and "Covid-19 pandemic" (Covid-19 pandemic). Significant cognitive and behavioral changes in users were triggered in specific contexts. These changes reflect users' responses to public health emergencies and reveal particular patterns of group behavior in extraordinary circumstances. Such unique behaviors and responses provide valuable insights into the psychology and actions of users when facing a crisis.

The third research direction focuses on communication modes, covering the three research fields of "populist communication," "driving factor," and "social media influencer." In the traditional media era, when information resources are relatively scarce and centralized, the channels for ordinary users to participate in public affairs are limited and scarce, and the communication process tends to be top-down, unidirectional, and static. Under the trend of equalization of information dissemination, fragmentation of information and integration of opinions have become the communication norm. The increased initiative of social media platforms has stimulated the active participation of online audience users, promoting interaction, discussion, and definition between different group labels. The desire of single individuals for autonomous expression drives the evolution and updating of group labels, reflecting the development of the communication model in the Internet era towards more dynamic, interactive, and pluralistic participation.

IV. Dissemination and Trends in Group Labeling

A. Generating the Environment

a) Scope of the group: Group labeling was initially a self-deprecating or biased identification behavior within online communities, and its dissemination was limited to relatively closed specific online communities and interest-based online community alliances. In the past decade, with the rise and development of social media platforms such as posting bars and microblogs, the user base of the platforms has been expanding, providing fertile soil for the generation and dissemination of group labeling, and group labeling, which was initially confined to a small circle, has begun to receive wider attention.

Especially in 2016, the emergence of short, fast, and downward social media platforms such as Shake, Shutter, Xiaohongshu, etc., with their user-friendly simple operation interface and personalized content recommendation algorithms, attracted a wide range of user groups from adolescents to older adults and greatly expanded the Scope of communication and influence of group labels [17]. The group tag communication mode has crossed into a new era, where group tags are no longer the exclusive entertainment of the creator and the referenced but have evolved into a communication element discussed and used by society in a universal scope. This shift from interest in specific groups to a pan-social craze signifies that group tagging has grown from a niche cultural phenomenon to a widely recognized mode of communication in society, and the expansion of its influence and dissemination also reflects the far-reaching impact and essential role of social media in modern society.

1) Label content: In recent years, the development trend of group labels has shown a more prominent characteristic of "depoliticization." Labels that used to directly reflect the social structure and carry critical colors, such as "second generation of rich people," "second generation of government officials," "second generation of demolition," etc., have been used less and less frequently. The ability of these labels to map the social structure directly is weakening. Public critical discourse has also weakened, and people have increasingly shifted their focus to their personal life status. The center of gravity of group labels has begun to carry more individual emotional expression and needs, such as emerging labels such as "Buddhist teenagers" and "melon eaters," which have become representatives of personal emotions and life experiences. These group labels reflect the fact that when people face marginalized situations, they autonomously choose a helpless "lying flat" mentality to cope with reality. To a certain extent, this phenomenon reflects the audience's reduced attention to the social environment and the gradual decrease in their participation in political life.

Regarding the vertical development of group labels, "depoliticization" has become an extremely significant trend. For example, group labels such as "kongjin" and "lemonjin" are descriptions of individual behaviors, while "laborer" and "985 waste" are personal creations of self-deprecating experiences. In this development process, the identity of individuality is becoming increasingly important. People look for resonance in labels that match their emotions and experiences and actively categorize themselves into various group labels given by society. Group labels are not only a reflection of social phenomena but also become a way for individuals to seek a sense of identity and belonging. Through these labels, people express their personalities and attitudes toward life while invisibly shaping society and culture's diversity. Although "depoliticized" labels may reduce direct criticism of social problems, they still convey, to some extent, individual perceptions and attitudes toward social reality.

2) Labeling behavior: With the frequent occurrence of online public opinion events, there is a significant phenomenon in the current fragmented communication environment: users increasingly rely on "labels" to define and understand people and things in the world around them. These "labels" are generalization tools to describe an object's salient features and attributes. Group labels share common characteristics by assigning individuals or behaviors to specific categories. Such group labels are not only an expression of collective attitudes and a depiction of group characteristics but also a redefinition of behavioral patterns and a marker of group identity and collective perception.

The popularization and increased Internet use have redefined the power structure of news and information publishing and dissemination. The dissemination of information is no longer the exclusive right of a minority class. However, it has become a universal right of the general public, which has also changed the process of constructing group labels. The low-threshold characteristics of the early Internet gave many Internet users the "right to speak," and the phenomenon of reverse labeling of the traditionally "underprivileged groups" for the "powerful groups" appeared [18]. This phenomenon has subverted the power structure of traditional society, allowing the general public to construct their group discourse and express their collective social attitudes and feelings.

From the social development process perspective, reverse labeling alleviates social conflicts to a certain extent and expresses resistance through online discourse banter. At the same time, the formation process of reverse labeling and the discussion within the group helps to unify internal attitudes and reduce the individual's dissatisfaction with the solidification of social class. This labeling process from the inside out helps group members to resonate with each other in terms of concepts and experiences, thus alleviating to a certain extent the sense of social imbalance brought about by class solidification.

With the diversified development of social platforms, the Internet is not only a channel for obtaining information but also a platform for emotional catharsis and expression of personal consciousness. In such a context, the construction of group labels is no longer to distinguish others. However, it has become a way for individuals to express their emotions, seek a sense of belonging, and realize their identity orientation through self-given behaviors. This self-labeling is an expression of the individual's social role and emotional state and a means for the individual to seek identity and positioning in society. In this way, individuals can find their position in the virtual space and realize their connection with the social group.

B. Dissemination Process

1) Initialization of dissemination: Information dissemination has several notable stages: spoken, written, printed, and electronic [19]. With the rapid development and successive iterations of Internet technology, online communication channels have proliferated, greatly enhancing the speed and Scope of information dissemination. At the same time, the increasing ease of use and accessibility of online social media platforms have enabled individual voices to reach a wide range of previously unimaginable audiences. In this process, the views expressed by individuals once resonated and supported by the group, can quickly converge into a powerful collective force. This power has subverted the one-way, closed pattern created by traditional media to a certain extent. Public social platforms constructed based on weak relationships have provided the general public with a place for free expression and communication, and they have promoted more authentic and frank self-expression and communication among the general public, providing individuals with new ways to express themselves and connect.

2) Spiral diffusion: WeChat provides users with a social media platform that is very different from open microblogs due to its private and robust relationship network. On this platform, the primary forms of socialization include private one-on-one conversations, group discussions, and content sharing and interaction in the circle of friends. Compared to public platforms such as Weibo, Douban, and Jieyin, WeChat's private nature provides users a safer and easier space for in-depth communication. Popular content on these platforms is often accompanied by a convenient sharing mechanism that allows users to easily forward information to WeChat or their circle of friends, thus realizing a seamless transition from public discussion to private communication. In the intensely relational space of WeChat, it is easier to form a common concern and emotional resonance for a specific topic because of the close social ties that usually exist between participants. Compared to the weak connections that may exist on public social platforms, interactions between individuals in the relational solid network on WeChat tend to be more frequent and in-depth, which not only strengthens the effect of information dissemination but also facilitates the exchange of emotions and the construction of a sense of identity among individuals under the group label.

3) Interpersonal communication: Some researchers and scholars believe that people rely on the information conveyed by the media to understand the environment and react based on it, which leads to the natural environment becoming closer and closer to the mimetic climate created by the media [20]. In this process, it is difficult for people to recognize the boundary between the mimetic and natural environments. This mimetic environment constructed by the media has a far-reaching impact on real society, which not only shapes the social order and the rules of interpersonal interaction but also continuously renews the dynamics of the social trend and has a significant social transformation function. Through its powerful influence, cyberspace's emerging information content and labeling continuously form new mimetic images in society. As a product of cyberculture, group labeling has begun to impact the natural world as its influence transcends the virtual space after extensive online use and dissemination. These labeled identities are brought into real-life interpersonal communication by the general public, gradually causing the natural environment to incorporate the characteristics of the online environment. This phenomenon shows that the popularity of group labels is not only spreading within virtual space but also penetrating and changing personal, identity, and social perceptions in the real world. In this way, group labeling has become an essential link between the virtual and the real, the individual and the society.

C. Evolution of Trends

In this paper, we selected the relevant literature on the emergence of group labeling during the ten years from 2013 to 2023, drew the keyword co-occurrence mapping according to CiteSpace 6.3 software, combined with the keyword clustering, and contacted the references to further analyze the vein of group labeling as a field of research in the context of online public opinion in Fig. 7.

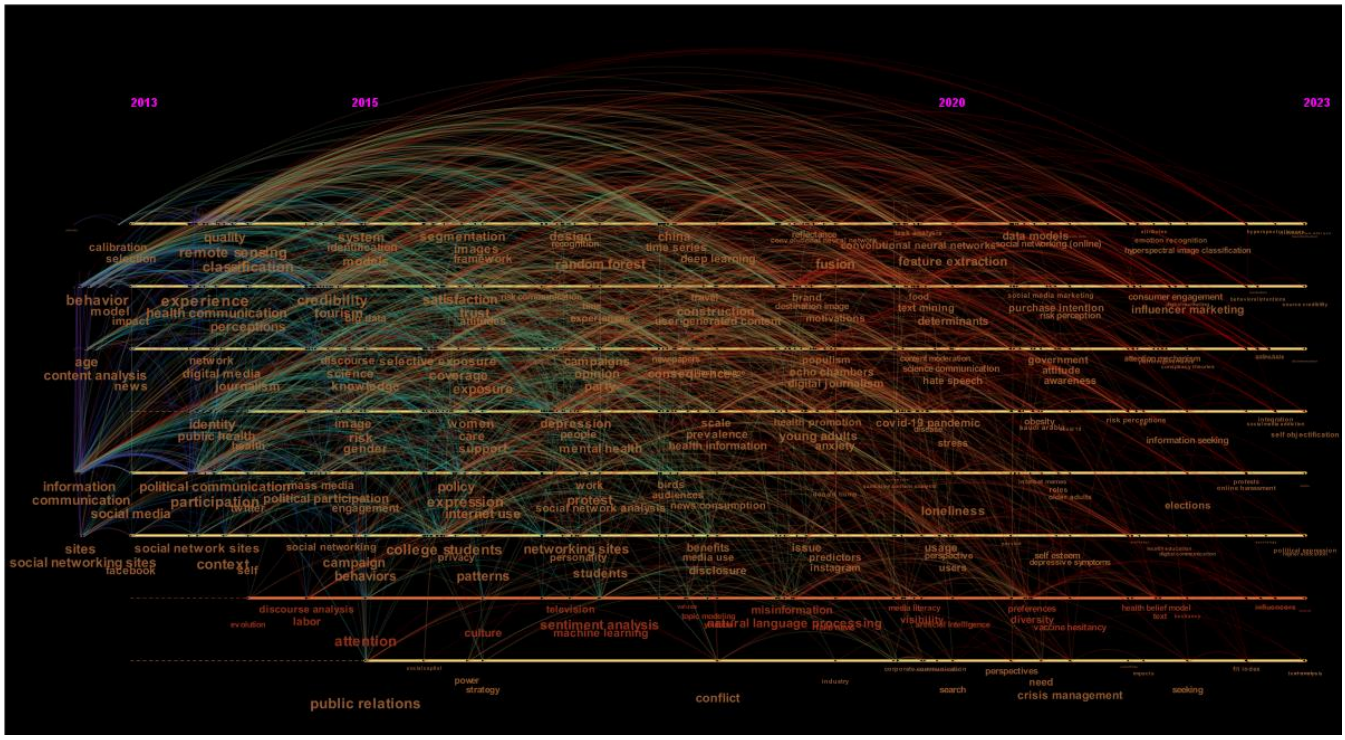


Fig. 7. Keyword temporal network mapping for group label propagation.

1) Exploration period: 2013-2015: The first stage is that research scholars mainly focus on Internet information technology, centering on keywords such as "action at a distance," "community," "information technology," "social media," and so on. Information technology (information technology)," "social media (social media)," and other keywords to explore the current Internet information technology to reshape the pattern of information dissemination. For example, the prosperity of microblogging, Douban, and other social platforms has changed the spatial distance between different media, the hierarchical level of similar media, and the distance between the communicator and the audience, which is enough to confirm that information dissemination in the age of the Internet has begun to develop in-depth from the mode of "flattening."

2) Exploration period: 2016-2020: In the study's second phase, research scholars focused on the development of group labeling on social media platforms and its impact. In this period, with the popularization of diverse social media platforms, every aspect of public life was infiltrated by online culture, and the rise of participatory culture brought about diverse cultural forms. Focusing on keywords such as "exposure," "machine learning," and "sentiment analysis," researchers have examined the subcultural nature of popular culture and group labeling in virtual space from a theoretical perspective. These studies explore how group labels are formed and developed in network culture and analyze how they reflect and influence social concepts and behavioral patterns. The idea of cultural capital plays a vital role in this process. It influences the expression and interaction of individuals on social media, making individuals gain more information and voice while at the same time controlling and changing the content and structure of group labels to some extent, contributing to the reduction of the depth and diversity of group labels due to the homogenizing influence of cultural capital.

3) Universal pan-tide: 2021-2023: Keywords such as "social media data," "awareness," "preferences," etc., have become hot topics for research scholars to focus on at present. Keywords such as "social media data," "awareness," "preferences," and so on have become the research hotspots that research scholars focus on at present. Comprehensively analyzing the above data, it is not difficult to find that group labels have become an effective way of globalized information dissemination and online public opinion dissemination relying on Internet technology and social media platforms, as well as an essential indicator of positive and negative energies and rationality in the online public opinion arena.

V. Conclusion

Our study has unveiled significant insights into the dynamics of online public opinion group labeling. Through a rigorous analysis of the data, we identified key trends and patterns that shape the evolution of group labels within the digital sphere. Our findings demonstrate that the diffusion of group labels is intricately linked to the interactive nature of social media platforms, which facilitates the rapid dissemination of labels and amplifies their impact on public discourse. Notably, our analysis revealed that the temporal dynamics of label adoption and the interplay between different user groups play a crucial role in the longevity and reach of these labels. These specific results not only advance our

understanding of online group labeling but also offer a foundation for future research aimed at developing more nuanced models of online communication.

Our study paves the way for further exploration in this domain. We recommend that future research delves into the application of alternative analytical tools, such as sentiment analysis and network theory, to gain a more comprehensive understanding of the emotional undercurrents and structural relationships within online communities. Additionally, exploring diverse datasets, including cross-cultural and cross-platform data, could provide a richer context for understanding the universality and specificity of group labeling phenomena. By extending our methodology to examine the role of group labels in shaping public opinion during significant events or crises, future studies can uncover the transformative power of these labels in real-world scenarios. These suggestions for future inquiry are not only aimed at refining our analytical approaches but also at broadening the scope of research to encompass the multifaceted influence of group labeling on society.

In this paper, the Web of Science Core Collection database is used as the data source of this study. Citespace 6.3 software is used to analyze the "group labeling." Through visualization and clustering analysis of the English literature co-citation, author co-citation, journal co-citation, keyword co-citation, clustering terms, etc., we analyze the generation, dissemination, and development trend of group labeling in online public opinion. Through visualization and cluster analysis, we analyzed the generation, dissemination, and development trend of group labels in the context of online public opinion.

ACKNOWLEDGMENT

This work supported by Soft Science Research Project of Henan Provincial Department of Science and Technology in 2024, *Research on the Communication Path of Red Culture in Colleges and Universities in the Era of Integrated Media*, No.: 242400410487.

REFERENCES

[1] Haider U A, Noman M, Ullah H, et al. An entirely passive ten-digit numeric keypad sensor using chipless RFID technology[J]. IEEE Sensors Journal, 2023, 23(3): 2978-2987.

[2] Noman M, Haider U A, Ullah H, et al. High-Capacity Double-Sided Square-Mesh-Type Chipless RFID Tags[J]. Electronics, 2023, 12(6): 1371.

[3] Anam H, Abbas S M, Collings I, et al. RFID Enabled Humidity Sensing and Traceability[C]//International Conference on Sensing Technology. Cham. Springer Nature Switzerland, 2022: 223-237.

[4] Suwalak R, Lertsakwimarn K, Lertwiriyaprapa T, et al. Dual-Band Band-Stop Filter using Multiple Hexagonal Microstrip Line for Chipless RFID Sensor [C]//2023 20th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI- CON). IEEE, 2023: 1-4.

[5] de Valle M K, Gallego-Garcia M, Williamson P, et al. Social media, body image, and the question of causation: meta-analyses of experimental and longitudinal evidence[J]. Body Image, 2021, 39: 276-292.

[6] Tylka T L, Rodgers R F, Calogero R M, et al. Integrating social media variables as predictors, mediators, and moderators within body image frameworks. Potential mechanisms of action to consider in future research[J]. Body Image, 2023, 44: 197-221.

[7] Rodgers R F, Rousseau A. Social media and body image: modulating effects of social identities and user characteristics[J]. Body Image, 2022, 41: 284-291.

[8] Pellegrino A, Stasi A, Bhatiasevi V. Research trends in social media addiction and problematic social media use: a bibliometric analysis[J]. Frontiers in psychiatry, 2022, 13: 1017506.

[9] Tiggemann M. Digital modification and body image on social media: Disclaimer labels, captions, hashtags, and comments[J]. Body Image, 2022, 41: 172-180.

[10] Harriger J A, Wick M R, Sherline C M, et al. The body positivity movement is not all that positive on TikTok: A content analysis of body positive TikTok videos[J]. Body image, 2023, 46: 256-264.

[11] Fardouly J, Slater A, Parnell J, et al. Can following body-positive or appearance-neutral Facebook pages improve young women's body image and mood? Testing novel social media micro-interventions[J]. Body Image, 2023, 44: 136-147.

[12] Prichard I, Taylor B, Tiggemann M. Comparing and self-objectifying: the effect of sexualized imagery posted by Instagram Influencers on women 's body image[J]. Body Image, 2023, 46: 347-355.

[13] McComb C A, Vanman E J, Tobin S J. A meta-analysis of the effects of social media exposure to upward comparison targets on self-evaluations and emotions[J]. Media Psychology, 2023, 26(5): 612-635.

[14] Wood-Barcalow N L, Alleva J M, Tylka T L. Revisiting positive body image to demonstrate how body neutrality is not new [J]. Body Image, 2024, 50: 101741.

[15] Powell J, Pring T. The impact of social media influencers on health outcomes: systematic review[J]. Social Science & Medicine, 2023: 116472.

[16] Dhadly P K, Kinnear A, Bodell L P. # BoPo: Does viewing body-positive TikTok content improve body satisfaction and mood?[J]. Eating Behaviors, 2023, 50: 101747.

[17] Becker E, Rodgers R F, Zimmerman E. # Body goals or# Bopo? Exposure to pregnancy and post-partum related social media images: effects on the body image and mood of women in the peri-pregnancy period[J]. Body Image, 2022, 42: 1-10.

[18] Rodgers R F, Gordon A R, Burke N L, et al. Parents and caregivers are critical players in the prevention and identification of body image concerns and eating disorders among early adolescents[J]. Eating Disorders, 2024: 1-24.

[19] Jarman H K, Fuller-Tyszkiewicz M, McLean S A, et al. Who is most at risk of poor body image? Identifying subgroups of adolescent social media users over a year[J]. Computers in Human Behavior, 2023, 147: 107823.

[20] de Valle M K, Wade T D. Targeting the link between social media and eating disorder risk: a randomized controlled pilot study[J]. International Journal of Eating Disorders, 2022, 55(8): 1066-1078.

Individual Cow Identification Using Non-Fixed Point-of-View Images and Deep Learning

Yordan Kalmukov[1], Boris Evstatiev[2], Seher Kadirova[3]
Department of Computer Systems and Technologies, University of Ruse "Angel Kanchev", Ruse, Bulgaria[1]
Department of Automatics and Electronics, University of Ruse "Angel Kanchev", Ruse, Bulgaria[2, 3]

***Abstract*—Monitoring and traceability are crucial for ensuring efficient and financially beneficial cattle breeding in contemporary animal husbandry. While most farmers rely mainly on ear tags, the development of computer vision and machine learning methods opened many new noninvasive opportunities for the identification, localization, and behavior recognition of cows. In this paper, a series of experimental analyses are presented aimed at investigating the possibility of identification of cows using non-fixed point-of-view images and deep learning. 14 objects were chosen and a photo session was made for each one, which provides training/validation images with different viewing angles of the animals. Next, a darknet-53-based convolutional neural network (CNN) was trained using YOLOv3, capable of identifying the investigated objects. The optimal model achieved 92.2% accuracy when photos of single or grouped non-overlapping animals were used. On the other hand, the trained CNN showed poor performance with group images, containing overlapping cows. The obtained results showed that cows could be reliably recognized using non-fixed point-of-view images, which is the main novelty of this study; however, certain limitations exist in the usage scenarios.**

Keywords*—Cow identification; convolutional neural network; YOLOv3; non-fixed point-of-view

I. Introduction

The advancement in information and communication technologies and artificial intelligence created numerous opportunities in all spheres of human society. They became the backbone of precision agriculture, allowing the optimization of all processes in the agricultural sector including animal husbandry. Nowadays, the food industry worldwide is oriented towards animal products, including meat, milk and milk products, eggs, etc.

An important aspect of livestock farming is the identification of animals, which offers many management and production benefits. Previously livestock owners were marking and branding their cattle with the main concern of proving their ownership. Many additional benefits are offered by contemporary animal identification, such as traceability of the herd and origin, tracking the cattle performance, disease management, keeping accurate track of the animal count, age, status, etc., and as a result, supporting the decision-making process.

The main approach in cattle identification is the application of ear tags. Each country has its regulatory framework to handle this problem, but in many cases, cows must be tagged after their birth. Furthermore, a cattle passport is created for them, including the date of birth, ear tag number, breed, sex, etc. [1]. Recently Radio Frequency Identification (RFID) and Electronic Identification (EID) tags are also popular, as they allow faster and automated data collection about the animals. Other approaches in animal identification are based on nose rings, collars, and image recognition [2,3,4,5]. Furthermore, many studies have suggested using multifunctional Internet of Things devices, which not only identify the animals but also allow monitoring and analyzing their behavior [2,6,7], tracking their location [8], etc.

When it comes to image processing for monitoring and recognition of cattle, the available studies investigate a wide range of opportunities, such as animal counting [9,10], individual cow recognition [11], behavior recognition and monitoring [12,13], animal tracking [14], cattle body detection [15], etc. Furthermore, in animal identification, so far only two image spectra have proven themselves appropriate for the above activities - visible (RGB) [6] and infrared [16].

The application of computer vision for the identification of species and animals has been investigated in numerous previous studies. In [16] the problems and opportunities when recognizing deer and wild boars based on infrared images and the You Only Look Once (YOLO) v3 neural network were studied. Similarly, in [17] RGB and infrared images obtained from camera traps, and a convolutional neural network (CNN), were used to identify frogs, lizards, and snakes. The recognition achieved an F1 score between 82% and 96% for the different species. This approach was also used in [18] with other wild animals, such as Amur tigers, Amur leopards, wild boards, several deer breeds, Asian black bears, red foxes, and other animals. The trained YOLOv5 neural network achieved precision and recall equal to 0.987 and 0.975, respectively.

Many other studies tried to identify not only the species but also to identify the specific animal. For example, in [19] RGB images and YOLOv7 were used for recognizing horses. The obtained mean average precisions (mAP) at the 50% threshold were 99.5% for identifying an animal when using its face and 99.7% when using its nose. Similarly, in [20] sheep faces were recognized based on RGB images. They were preprocessed using noise removal, brightness/saturation/contrast adjustment and when necessary horizontal flipping, and were thereafter fed to different neural networks. The highest accuracy and F1 score were achieved with RepB-Sheepnet, reaching more than 99%. Face recognition for identifying cows with a dual 3D camera setup and the Iterative Closest Point method was used in [21]. The authors have chosen to use the face of the cows, because of the rigid structure of the skull. The reported

identification rates varied between 88% and 99%, depending on the gallery point clouds per cow. In another study, the Siamese DB Capsule neural network was applied for the face recognition of cattle [22]. The used approach pairs the images, i.e. each image is compared with another one, and "pos" or "neg" is returned, depending on whether the images belong to different or the same categories. The study achieved recognition accuracy of up to 93% and an F1 score of up to 93.54%. Several other studies have also tried to identify cattle by their noses. In [23] cattle were recognized using their noses using RGB images as row data. The proposed methodology includes supplying grayscale images to a Deep belief network DBN for the recognition process, which was implemented using the Matlab R2019b toolkit. The authors reported that the recognition accuracy reached up to 99% with 400 training images, and it decreases when fewer images are used. Nose-based cattle identification was also used in [24], where the muzzle patterns of each animal were obtained with the help of numerous deep-learning algorithms. The highest achieved accuracy was 98.7% for a VGG16_BN-based model.

The recognition of the cows can also be classified depending on where the cameras are positioned and how many cameras are used. Common approaches are using topview, sideview, and backview cameras. In [25] the cows were photographed from above, with the idea of recognizing their skin pattern. The methodology includes background removal, image rotation, alignment according to the template, and pixelized binary image creation. The nearest neighbor approach and the Hammering distance measure were used for the recognition process. The authors reported a Top-1 accuracy of 61.5% and a Top-4 accuracy of 83%. Furthermore, they stated that such an approach does not require retraining the recognition algorithm (compared to deep learning) and is very fast. A similar approach was used in [26], where photographs of the cows' bodies and a neural network were used to recognize their body patterns. The reported accuracy was more than 92% for the training data and 90% for the testing data.

In [27] 7069 topview images of 62 cows were used for training, 1801 images for validation, and 1104 images for testing different CNNs. The ResSTN model achieved the highest average recognition accuracy of 94.58% and slightly higher accuracy under daylight lighting conditions. In [28] daytime and nighttime topview images were used to track the cattle movement over a farm. The idea was to use many strategically placed cameras not only to identify the animal but also to classify its behavior into the categories "resting", "standing", "standing up" and "walking". Different versions of YOLOv5 were used to identify the animals, which achieved mAP ranging from 92.7% to 95.3%.

Other studies used sideview cameras for cattle identification. In [29] RGB images of 13 cows were used for their recognition using a convolutional neural network with ResNet50 as the backbone. The authors reported more than 98% recognition accuracy of the investigated objects. Similarly, in [30], side RGB images of cows were used for their recognition. Different color spaces were used to obtain the most appropriate one for distinguishing the animals. The Euclidean distance of feature vectors of critical points was used as the criteria for identification, together with the Brute Force Matcher algorithm. The optimal accuracy of 99.31% was obtained for the Lαβ color component. Another approach for the identification of animals was presented in [31] that relies on rump RGB images. 2140 images of 195 cows were used for training a convolutional neural network and 917 for its validation. The CNN based on a Mobilenet v2 backbone returned the highest accuracy, reaching up to 99.76%.

In [32] three cameras were used for cow recognition - topview, frontview and sideview. An enhanced filter algorithm was proposed, combining the mean-shift and particle-Kalman filter algorithms. The image processing was implemented using Matlab, though no accuracy has been reported. Several cameras were also used in [33], where top and sideview RGB and depth images from a 15-fps video were applied for cow identification. The study uses Euclidean cluster extraction to select the largest 3D point cluster representing the cow, and then to estimate the average silhouette of the animal. Thereafter, the differences between the obtained silhouette and the probed one are evaluated. The achieved algorithm accuracy reached 75.6%.

A completely different approach was used in [34]. The YOLOv3 neural network was utilized to "read" the ear tag of cows, while they were near the drinker. The idea was to estimate how long the animals are drinking and as a result to estimate the approximate volume of water drunk. A mean average precision of 89% and an F1 score of 86% were obtained.

Other studies had more advanced goals, such as behavior identification. In [30] the calving time of cows was predicted via motion classification using a 360° overhead RGB camera. The methodology includes object identification, background subtraction, generation of the object contour, and principal component analysis (PCA) to extract features. An average accuracy of up to 95% was reported for detecting and classifying cow motions.

The analysis of previous animal recognition studies showed that deep learning provides acceptable results. Different barriers exist when using deep learning for cow identification, such as the limited number of images of each object, the fluctuations of the positions and angles of view, the appearance of numerous cows the network wasn't trained for, etc. [22]. Furthermore, often it is hard for a human being to tell the difference between two separate cows, especially if they are single-colored, which might be a problem when preparing the training/validation data. Most previous studies have used images of the animals, taken from a specific point of view (side, back, frontal, etc.) or from above that are rotated to an appropriate orientation. However, there are almost no studies, dealing with the identification of cows with non-fixed point-of-views. The abovementioned shows the existence of a knowledge gap in this area, which should be addressed.

This study aims to investigate the possibility of real-time identification of cows using deep learning and images, which were obtained without limitations for the angle of view and orientation of the animals. To achieve this the neural network should be trained with numerous images, representing the animals from different points of view.

II. Materials and Methods

A. The Study Area

The experimental part of this study was conducted at a farm for outdoor cow breeding. It is located in the village of Trastenik, Ruse District, north-central region of Bulgaria (Fig. 1) and has coordinates 43.65830753698392, 25.845059022235304. 30 dairy cows from the Bulgarian Black and White cow breed and the Red and White Holstein cow breed are bred on the farm. All images and videos of cows were shot on a summer day (5 July 2023), late in the afternoon between 5 and 7 p.m.

Fig. 1. Location of the experimental pasture.

B. Methodology for Data Collection and Data Processing

The data collection and processing methodology, applied in this study is summarized in Fig. 2. It can be divided into five main steps, which are explained below.

1) Step 1. Data collection: This step begins with choosing the cows that are the object of the investigation; thereafter, numerous images of each object are made. The successful recognition of the object requires enough images representing the animal from different sides and in different circumstances (staying, grazing, etc.). This is achieved in two ways:

- By making numerous photos of the animal;
- By filming a video of the animal from all sides and extracting appropriate frames from it.

The described procedure is repeated for each cow that should be investigated. Taking photos from different distances and with different backgrounds should also be considered.

2) Step 2. Data selection: In this phase, the already collected data is analyzed, sorted, and filtered. Initially, all photos taken that contain a certain object are sorted in separate folders (per object). Furthermore, if a video was made of a certain animal, frames are extracted from it as images, representing the animal with different viewing angles. They are also sorted in the corresponding folders. Finally, several of the prepared training images are filtered out for the validation set so that they are not used during training.

3) Step 3. Data preparation: The goal of this step is to prepare the training and validation data. Each object is marked in a rectangle and classified using LabelImg or an alternative tool. This is repeated for each of the images and each of the objects. If a certain image contains more than one of the investigated objects, all of them could be marked and categorized in a different class, as shown in Fig. 3. In this step both the training and the validation images are classified.

4) Step 4. Deep learning: In the next data processing phase a machine learning model is trained. In this study, we use the YOLO v.3 object recognition system, and therefore all data should be prepared and sorted accordingly [35]. We chose version 3 because it has shown good results in previous studies. The preparation for this step includes:

- tuning up a config file corresponding to the graphical processing unit (GPU) characteristics, the number of classes, and the number of training iterations;
- setting up the training itself, i.e. selecting the training and testing data, the config file, the initial weights as well as the application of mean average precision (mAP).

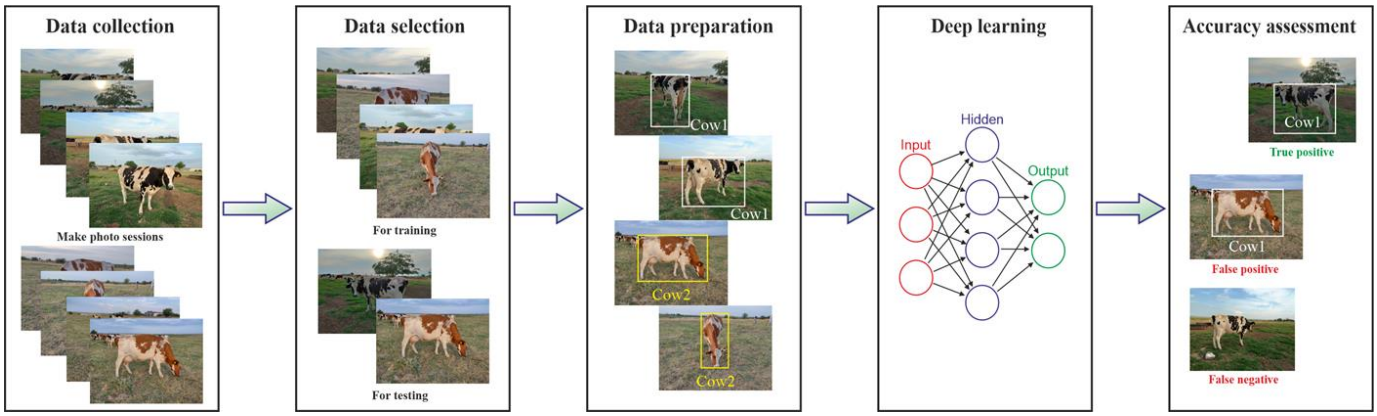


Fig. 2. Overview of the used methodology.

Fig. 3. Marking each object with a rectangle in the LabelImg tool.

5) Step 5. Accuracy assessment: The final step of the data processing is to estimate and evaluate the models' accuracies. In this study this is implemented from several perspectives:

- The first one is the automatic accuracy assessment during the training of the models, which is implemented by YOLO itself and is based on the mean average precision performance metric. It is achieved using validation images prepared in steps 2 and 3. The results from this assessment are taken into account when selecting the optimal classification model, i.e. the one with the highest mAP. The meaning of mAP metric is as follows:

$$mAP = \frac{1}{N}\sum_{n=1}^{N} AP(n) \quad (1)$$

where N is the total number of classes and AP(n) is the average precision for a given class n, which is calculated as the weighted mean of precision at each threshold.

- The second perspective is to use additional images and videos, which were not used for training and validation, and a human operator to confirm that the recognized objects (cows) are correct or incorrect. For images, this evaluation is straightforward. Videos could be processed in the following way:
  - Each video is resampled to 1 frame per second (FPS) framerate.
  - The video is analyzed with YOLO v.3 using the selected optimal model and the analysis results are saved as video files using screen recording software like OBS Studio or alternative.
  - The recorded video is observed by an operator, frame by frame, and a confusion matrix is created for each object.
  - The model performance is assessed in terms of Precision, Recall, and F1 Score, whose meaning is described below. The Precision metric assesses the accuracy of positive predictions and is defined with:

$$P = \frac{TP}{TP+FP} \quad (2)$$

The Recall metric gives the proportion of true positive (TP) predictions among all positives and is defined as:

$$R = \frac{TP}{TP+FN} \quad (3)$$

The F1 score metric balances the score of precision and recall according to:

$$F1 = 2 \times \frac{P \times R}{P+R} \quad (4)$$

Finally, the average accuracy for the whole testing dataset is estimated using the following equation:

$$Accuracy = \frac{Number\ of\ testing\ frames}{Number\ of\ testing\ frames+FP+FN} \quad (5)$$

Next, the obtained results are evaluated and analyzed using the created confusion matrix. Situations with false positives (FP) and false negatives (FN) are closely analyzed, to identify the reasons behind this.

III. Results and Discussion

A. Training of Convolutional Neural Networks

801 images of 14 different dairy cows were used for training. Some of the cows look very similar and are quite difficult to distinguish even for a human being (Fig. 4). For most cows 50-70 photos each were used, though there are also cows photographed over 100 times, while others were shot far less than that (just 10-15 images). All photos were taken under natural environment conditions, without separating the animals from the herd or placing them in an enclosure. An Mx-

M16TB-R079 RGB camera by Mobotix AG (Langmeil, Germany) in continuous video recording mode and three mobile phones were used for filming. The resulting videos and images have different resolutions and color saturation.

It should be noted that most of the cows look different from their two sides, which makes it important to select an appropriate collection of photos, representing the animals from all sides for training an adequate model. That is why for each cow an improvised photo session was made (Fig. 5).

Next, according to the methodology, a folder with images was made for each object. Furthermore, frames from the video recordings were chosen, extracted, and placed in the corresponding folders. On each image, the target objects were selected using the LabelImg tool. Thereafter, the images were divided for training, validation, and testing purposes, as shown in Table I. A total of 44 images were used for validation and 37 images of individual cows for testing. The table data shows that the datasets are imbalanced, which should be considered when interpreting the results.

The training of the CNN was performed using an NVidia RTX 3060 GPU with 12 GB dedicated video RAM, which allowed roughly 1000 training iterations per hour. According to the developers of YOLO, the recommended number of training iterations is the number of classes multiplied by 2000, but not less than the number of images. Therefore, initial training was conducted with a maximum number of 28000 iterations. With such a configuration, YOLO saves the calculated weights at every 10000 iterations, at the maximum number of iterations (28000), and at the maximum mAP value. The latter is calculated during the training process using the validation dataset. The highest obtained mAP is 100% and was achieved at 18500 iterations.

In previous training sessions with cow images, we have observed that the neural network could be easily overtrained and begin to miss (fail to recognize) cows that it was trained for, but which were viewed from a slightly different position than those in the training dataset. Experiments with different numbers of training iterations showed that more cows (especially in group photos) were recognized with weights obtained in fewer iterations. To avoid overtraining, we retrained the network with 9000 iterations, thus the YOLO system saved the weights every 1000 iterations. As a result, we obtained multiple alternative weight files that allow us to perform additional experimental analyses and determine the optimal number of iterations. When training with 9000 iterations, the maximum value for mAP is 98.81% and was reached at 6850 iterations. The training and validation results with 9000 iterations are shown in Fig. 6.

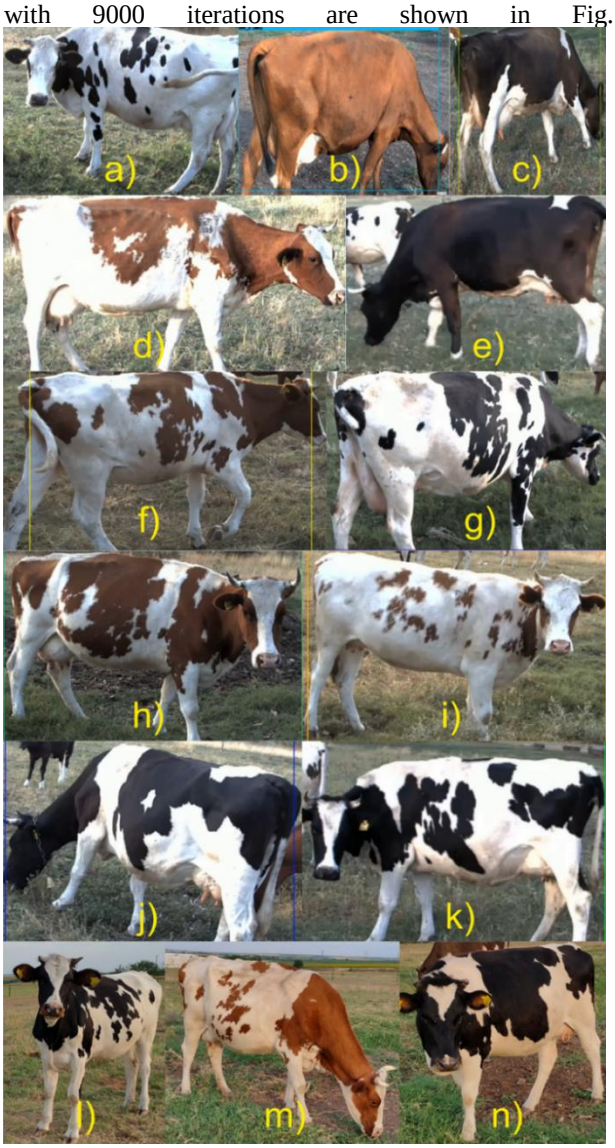


Fig. 4. The investigated cows.

After a series of experimental analyses, we found that for both group photos and videos, more cows are recognized when using the weights file obtained at 6850 iterations, which is why it is used for all subsequent experiments.

Fig. 5. Taking a photo session of cow №8 from different points of view: a) right side; b) back side; c) left side; d) front side.

TABLE I. SUMMARY OF THE NUMBER OF IMAGES OF EACH OBJECT USED FOR TRAINING THE CONVOLUTIONAL NEURAL NETWORK

Cow №	1		2	3	4	5	6	7	8	9	10	11	12	13	14
Training images	52		24	52	149	80	35	71	81	58	72	49	11	45	22
Validation images	3		1	2	4	4	2	4	3	5	5	4	1	4	2
Testing images	3		2	3	7	2	1	2	5	3	2	2	1	2	2
Testing video frames	89		117	166	139	187	58	93	91	111	165	232	/	/	/

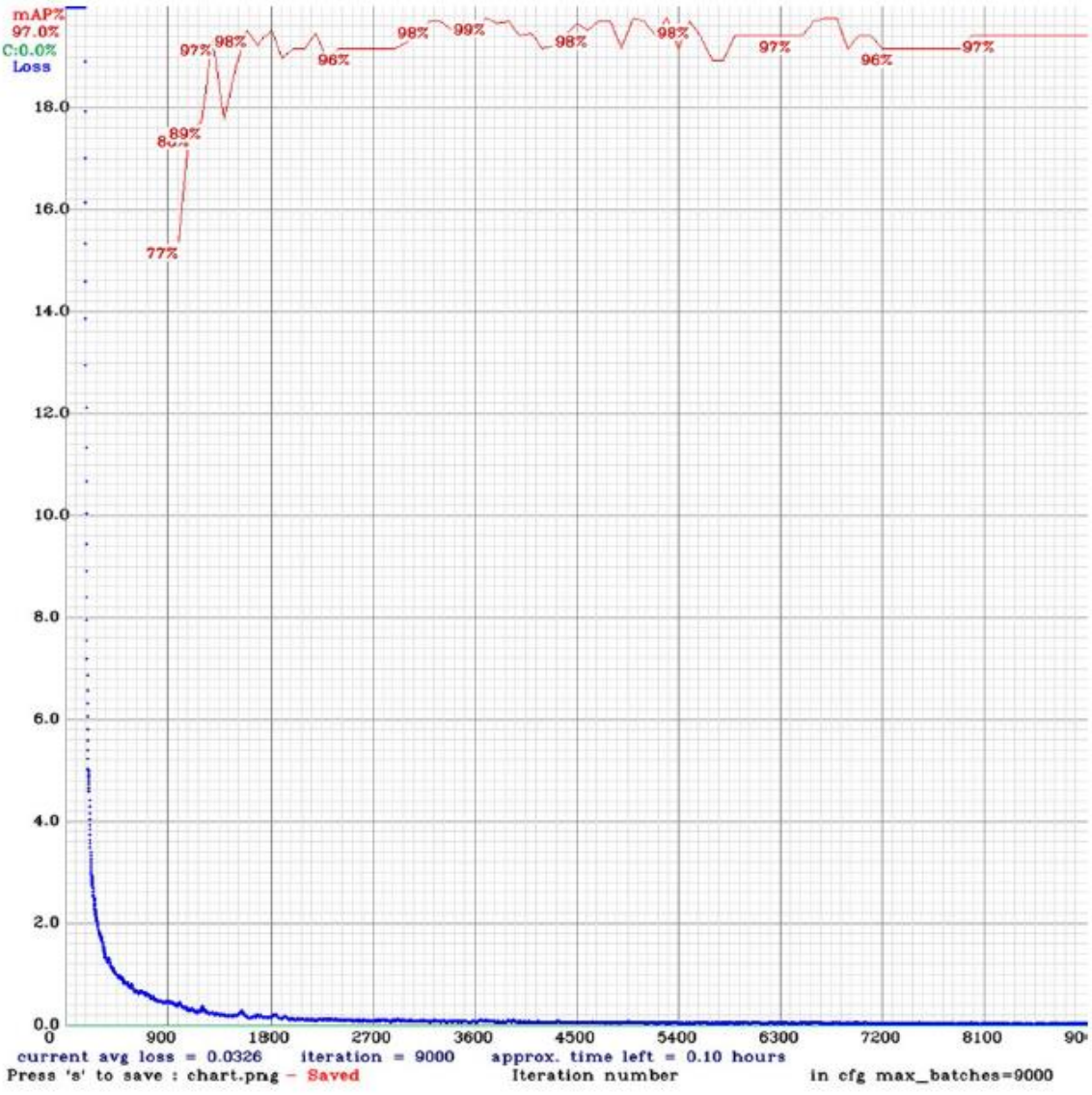


Fig. 6. YOLO training results with 801 images of 14 cows at 9000 iterations, with maximum mAP 98.81% at 6850 iterations.

It is interesting to note that at 18500 iterations mAP has a maximum value of 100% and cows in individual pictures are recognized with a maximum probability of 0.99 or 1. At 6850 iterations, despite the slightly lower mAP value of 98.81% and lower individual probabilities, more animals are recognized in group photos and videos. This suggests that at 18500 iterations, the neural network is already overtrained. Another possibility is that the validation dataset, which is 5.625% of the training dataset, is too small and could be increased in future studies.

B. Assessment of the Accuracy

Initially, 37 testing images not applied in the training and validation process were used. Their distribution between the different objects of the study is shown in Table I. In all of them, only one cow is visible or the other cows are positioned behind the recognized object. In this scenario, the achieved recognition rate is 100% and this applies to all 14 investigated objects. Furthermore, the recognition is done with a probability between 98% and 100%. Fig. 7 presents several examples of the correctly recognized cows from the testing dataset.

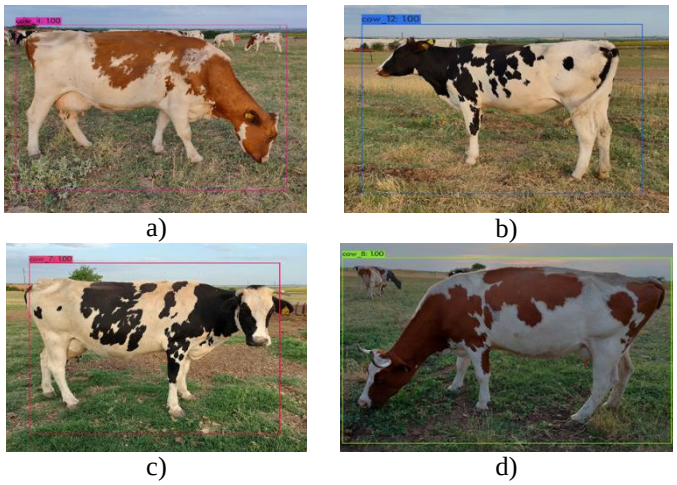


Fig. 7. Examples of recognized cows with high (typically 100%) probability in cases where the images contain only one cow and the animal is seen in sufficient detail.

A significantly different situation occurs when there are numerous cows in the image. In some cases, the model was able to recognize correctly two cows in a single image, as shown in Fig. 8. Nevertheless, it could be noticed that in Fig. 8, a cow_8 is identified with low probability (33%), which might be explained by the fact that it is partially visible. Similarly, in Fig. 8(b) again cow_8 is recognized with a probability of 31%, which might be caused by its overlapping with other similarly colored cows.

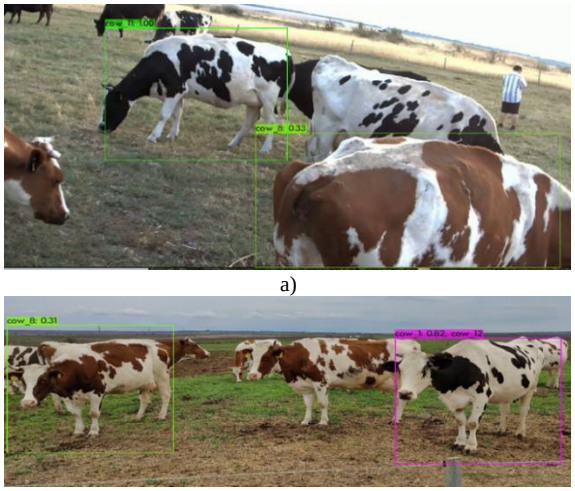


Fig. 8. Examples of correctly recognized two cows in a single image: a) one of the objects is partially visible; b) both objects are fully visible and overlapping with other cows.

However, that is not always the case. In many other situations, the trained model experienced difficulties recognizing multiple cows in group photos, although it was trained for many of them. In these cases, usually up to 1 and more rarely two cows in the front are recognized using the weights obtained at 6850 iterations. If the weights file, obtained at 18 500 iterations is used instead, the results are worse. This problem is demonstrated in Fig. 9, where cows numbered five and eight are available in the two photos, which were taken with a time difference of several seconds. Nevertheless, the trained model identifies either one or the other, but not both objects. In this example the probability rate is relatively high – 70% and 88%, respectively. It could be noticed that in Fig. 9(a) there are other cows behind cow_5, which might be influencing its recognition performance. Similarly, in Fig. 9(b) cow_8 is partially overlapping with cow_5, which once again might be the reason for such behavior.

More problems with the identification of cows in group photos are demonstrated in Fig. 10 and Fig. 11. The first one shows that two objects (cow_5 and cow_4) were not identified. The most probable reason is that the model fails to distinguish them as separate objects from the numerous surrounding animals. Cow_5 overlaps with two other cows and cow_4 with one in front and many others in the background. This suggestion is confirmed by the example, demonstrated in Fig. 11, where three cows are identified as a single animal (cow_1), even though only one of them is cow_1.

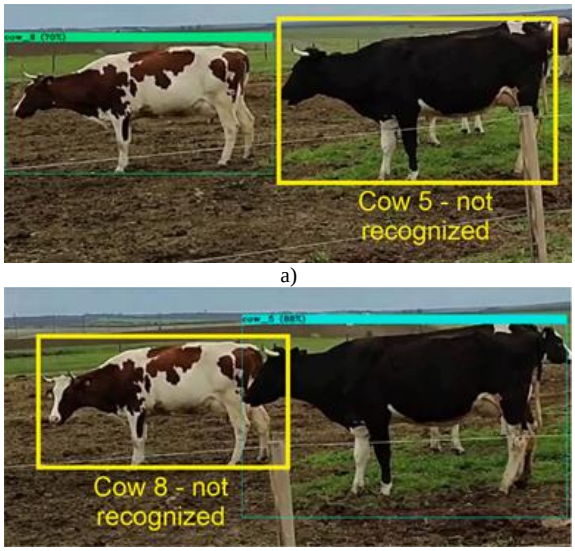


Fig. 9. An example of recognizing cow 8 (a) and cow 5 (b) in two photos of the two objects.

The reason behind the abovementioned problems might be the used datasets. In fact, no group images of cows were included on purpose in the training/validation datasets, which might have limited the identification ability of the CNN model under such circumstances.

Recognizing cows in videos is very similar to recognizing objects in still images, just the video should be preprocessed and, in most cases, converted to multiple still images (shots). In our case, the video of each cow was analyzed and converted to a video with 1 fps framerate. Thereafter, an operator manually classified each frame of the video as either true positive, false negative, or false positive to assess the accuracy of the CNN model. This way the performance of the model is evaluated individually for each object. In the following experimental analysis, we assessed the model performance only for the first 11 cows since video materials were only available for them. The total number of video frames used for testing is 1448.

Each object was analyzed from two perspectives:

Fig. 10. An example of 2 cows not recognized in a group photo with many others overlapping.

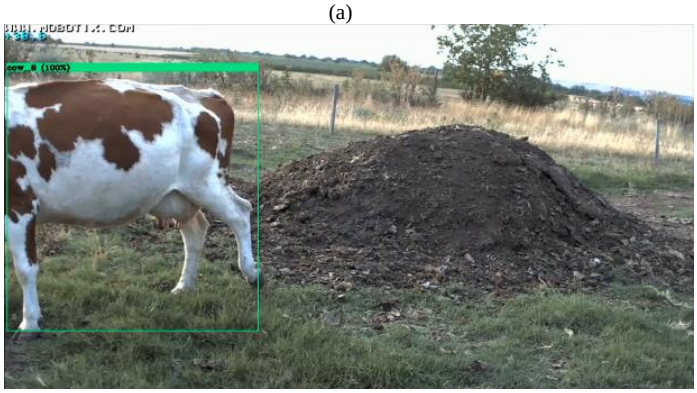
Fig. 11. An example of 3 overlapping cows recognized as a single object.

- The animal is fully visible by the camera [Fig. 12(a)] – it is assumed that at least the following is visible: full carcass, full head (unless it is behind the carcass), and at least half of the legs.
- The animal is only partially visible by the camera [Fig. 12(b)] – at least one of the abovementioned requirements is not met. Nevertheless, large enough part of the cow should be visible, so that the operator can recognize it.

The results from the analysis of the video frames, including true positives, false negatives, false positives, precision, recall, and F1 score are summarized in Table II. When the objects are fully visible the F1 score varies between 0.89 and 1.00 for the different objects and is above 0.95 for most of them. These results are very similar to those obtained in still images of single animals, which corresponds to the expectations. The only exceptions are cow 2 (0.89) and cow 6 (0.90). Yet, this is not unexpected because a limited number of images for these objects were used to train the CNN model - 24 and 35, respectively; while for all other objects, the available images are more than 50. The average accuracy of the trained optimal model was obtained to be 92.2% when used with non-group photos.

When the cows are not fully visible, the F1 score varies between 0.43 and 0.98 for the different objects. The lowest results were achieved for cow_6 (0.43), cow_5 (0.80), and cow_11 (0.84), which is caused mainly by the lower recall, which accounts for the influence of the false negatives. In most cases, the cropped animals were simply not recognized, which is an expected behavior in such situations.

(a)



(b)

Fig. 12. Examples for fully visible (a) and partially visible (b) cow, according to the accepted rules in this study.

TABLE II. A SAMPLE CONFUSION MATRIX

Situation	TP[c]	FN[d]	FP[e]	Precision	Recall	F1
Cow 1 (fully visible) [a]	31	0	2	0.94	1.00	0.97
Cow 1 (partially visible) [b]	40	8	0	1.00	0.83	0.91
Cow 2 (fully visible)	31	8	0	1.00	0.80	0.89
Cow 2 (partially visible)	58	3	0	1.00	0.95	0.98
Cow 3 (fully visible)	71	0	2	0.97	1.00	0.99
Cow 3 (partially visible)	57	4	2	0.97	0.93	0.95
Cow 4 (fully visible)	98	0	0	1.00	1.00	1.00
Cow 4 (partially visible)	26	5	2	0.93	0.84	0.88
Cow 5 (fully visible)	112	2	11	0.91	0.98	0.95
Cow 5 (partially visible)	37	5	13	0.74	0.88	0.80
Cow 6 (fully visible)	42	2	7	0.86	0.95	0.90
Cow 6 (partially visible)	3	6	2	0.60	0.33	0.43
Cow 7 (fully visible)	34	0	0	1.00	1.00	1.00
Cow 7 (partially visible)	30	9	1	0.97	0.77	0.86
Cow 8 (fully visible)	63	0	0	1.00	1.00	1.00
Cow 8 (partially visible)	22	1	0	1.00	0.96	0.98
Cow 9 (fully visible)	68	0	0	1.00	1.00	1.00
Cow 9 (partially visible)	35	1	1	0.97	0.97	0.97
Cow 10 (fully visible)	126	1	2	0.98	0.99	0.99
Cow 10 (partially visible)	24	0	4	0.86	1.00	0.92
Cow 11 (fully visible)	126	5	2	0.98	0.96	0.97
Cow 11 (partially visible)	28	11	0	1.00	0.72	0.84
Average (fully visible)	802	18	26	0.97	0.98	0.97
Average (partially visible)	360	53	25	0.94	0.87	0.90

[a.] The following parts of the animal are visible: the whole carcass, at least half of the legs, and the whole head. If part of the carcass or head are not visible because of the angle of view, the cow is also considered to be fully visible.

[b.] In all other cases the cow is considered to be only partially visible.

[c.] The object was correctly recognized

[d.] The object was not recognized or was incorrectly recognized.

[e.] Another object was recognized as this one.

C. Comparison with Previous Studies

The results obtained in this study cannot be directly compared to other studies, as to the best of our knowledge no previous studies have tried to identify cows using mixed images, showing them from different points of view. Nevertheless, we can compare our results with those obtained in studies, using topview, sideview, backview and faceview images, independently. In [27] outside topview images of cows were used, representing either the full body or randomly cropped body. The obtained accuracy was 95.23% and 90.85%, respectively, which generally corresponds to the average 92% accuracy, achieved in our study.

In [29] sideview full body images were used to achieve 98.58% accuracy. Other similar studies are [36] and [37], reaching average accuracies of 96.65% and 90.2%, respectively. In [37] an F1 score of 86% was estimated, which is lower than the one obtained in this study (97% and 90%, respectively for full body and cropped body). In [31] backside full body images of cows were used to achieve 99.76% accuracy, and in [22] head images for face recognition were used with 93% accuracy and 93.54% F1 score.

Table III summarizes the results from the performed comparison. It can be noticed that most studies assessed only the accuracy of the trained models, which is known to be misleading in the case of imbalanced datasets. That is why in our study we have also obtained precision, recall, and their average F1 score, which gives a more accurate evaluation of the performance of the trained CNN.

TABLE III. COMPARISON OF OUR RESULTS WITH THOSE, OBTAINED IN PREVIOUS STUDIES

Paper	CNN backbone	View	Location	Cropping	Accuracy	F1 score
Wang et al. [27]	ResSTN	Topview	Outside	Full body	95.23%	N/A
				Randomly cropped	90.85%	N/A
Fu et al. [29]	ResNet50	Sideview	Inside	Full body	98.58%	N/A
Hou et al. [31]	Mobilenet v2	Backside	Inside	Full body	99.76%	N/A
Shen et al. [36]	AlexNet	Sideview	Inside	Full body	96.65%	N/A
Wang et al. [37]	ResNet101	Sideview	Outside	Full body	90.2%	86%
Xu et al. [22]	Capsule network	Faceview	Inside	Head only	93.00%	93.54%
Ours	Darknet-53	All sides	Outside	Full body	92.2%	97%
				Randomly cropped		90%

IV. CONCLUSIONS

In this study, the possibility of training a convolutional neural network for recognizing cows regardless of the viewing angle was investigated. A dataset containing 801 images of 14 cows was used for training a YOLOv3 model. To prepare the training, validation, and testing datasets, photo sessions of each animal were made, so that they represent the cows from different sides. Out of the trained models, the one selected as optimal achieved a mean average precision of 98.81% after 6850 iterations.

The results from the evaluation of the trained CNN showed that its recognition rate greatly depends on the usage

circumstances. The model showed excellent performance at recognizing objects on images/video frames, where there are no other animals, where the animal is in the front, or where the animal is not overlapping with other animals. This is confirmed by the obtained 97% F1 score for fully visible cows and 90% for partially visible cows, as well as by the achieved average accuracy of 92.2%. On the other hand, the trained YOLOv3 CNN does not perform well when numerous objects exist on the image/video frame, especially with overlapping objects. In such a situation, the model fails to identify all cows and rarely recognizes more than one.

The abovementioned indicates that the proposed approach for the recognition of cows with a non-fixed point-of-view is applicable but with certain limitations and under certain usage scenarios. Such limitations are: the object being identified should be alone on the image/frame, or should not be overlapping with other animals. The investigated approach has numerous possible applications, such as monitoring the animals' movement, localization, and behavior, or even as a replacement for ear tags.

In the present study, 801 images for training were used, which is 57 on average per animal. It is interesting to investigate if the performance of CNN models will increase when more training images are used, especially in group photos with overlapping cows. Furthermore, more group photos of animals should be included in the training/validation datasets, which might increase the recognition rate in such scenarios. If that doesn't happen, another option is to train an additional neural network, responsible for recognizing the extents of each object. With such an approach each cow could be extracted from the image and identified independently of the others, which should allow reducing the impact of overlapping animals. The abovementioned represent promising research topics, which are an object for future studies.

ACKNOWLEDGMENT

This research was funded by the Ministry of Education and Science of Bulgaria under the National Research Program "Intelligent Animal Husbandry", grant number Д01-62/18.03.2021.

REFERENCES

[1] M. Knock, G.A. Carroll, "The Potential of Post-Mortem Carcass Assessments in Reflecting the Welfare of Beef and Dairy Cattle," Animals, vol. 9, no. 959, 2019, doi: https://doi.org/10.3390/ani9110959

[2] P. Kumar, S. Luo, and K. Shaukat, "A Comprehensive Review of Deep Learning Approaches for Animal Detection on Video Data" International Journal of Advanced Computer Science and Applications(IJACSA), vol. 14, no. 11, 2023, doi: http://dx.doi.org/10.14569/IJACSA.2023.01411144

[3] T. Gao, D. Fan, H. Wu, X. Chen, S. Song, Y. Sun, J. Tian, "Research on the Vision-Based Dairy Cow Ear Tag Recognition Method," Sensors, vol. 24, no. 2194, 2024, doi: https://doi.org/10.3390/s24072194

[4] J. Li, Y. Liu, W. Zheng, X. Chen, Y. Ma, L. Guo, "Monitoring Cattle Ruminating Behavior Based on an Improved Keypoint Detection Model," Animals, vol. 14, no. 1791, 2024, doi: https://doi.org/10.3390/ani14121791

[5] Y. Kalmukov, B. Evstatiev, "Methods for Automated Remote Sensing and Counting of Animals," 2022 8th International Conference on Energy Efficiency and Agricultural Engineering (EE&AE), pp. 1-7, IEEE, 2022, doi: https://doi.org/10.1109/EEAE53789.2022.9831239

[6] S. Hu, A. Reverter, R. Arablouei, G. Bishop-Hurley, J. McNally, F. Alvarenga, A. Ingham, "Analyzing Cattle Activity Patterns with Ear Tag Accelerometer Data," Animals, vol. 14, no. 301, 2024, doi: https://doi.org/10.3390/ani14020301

[7] T. Mladenova, I. Valova, B. Evstatiev, N. Valov, I. Varlyakov, T. Markov, S. Stoycheva, L. Mondeshka, N. Markov, "Evaluation of the Efficiency of Machine Learning Algorithms for Identification of Cattle Behavior Using Accelerometer and Gyroscope Data," AgriEngineering, vol. 6, no. 3, pp. 2179-2197, 2024, doi: https://doi.org/10.3390/agriengineering6030128

[8] T.T. Zin, M.Z. Pwint, P.T. Seint, S. Thant, S. Misawa, K. Sumi, K, Yoshida, "Automatic Cow Location Tracking System Using Ear Tag Visual Analysis," Sensors, vol. 20, no. 3564, 2020, doi: https://doi.org/10.3390/s20123564

[9] K. Arai, T. Higuchi and T. Murakami, "Estimation Method of the Total Number of Wild Animals based on Modified Jolly's Method" International Journal of Advanced Computer Science and Applications(IJACSA), vol. 8, no. 1, 2017, doi: http://dx.doi.org/10.14569/IJACSA.2017.080143

[10] O. Ittoo and S. Pudaruth, "Automatic Recognition of Marine Creatures using Deep Learning" International Journal of Advanced Computer Science and Applications(IJACSA), vol. 15, no. 1, 2024, doi: http://dx.doi.org/10.14569/IJACSA.2024.0150106

[11] Z. Weng, S. Liu, Z. Zheng, Y. Zhang, C. Gong, "Cattle Facial Matching Recognition Algorithm Based on Multi-View Feature Fusion," Electronics, vol. 12, no. 156, 2023, doi: https://doi.org/10.3390/electronics12010156

[12] N. M. Arago, C. I. Alvarez, A. G. Mabale, C. G. Legista, N. E. Repiso, R. R. A. Robles, T. M. Amado, R. Jr. L. Jorda, A. C. Thio-ac, J. S. Velasco, and L. K. S. Tolentino, "Automated Estrus Detection for Dairy Cattle through Neural Networks and Bounding Box Corner Analysis" International Journal of Advanced Computer Science and Applications(IJACSA), vol. 11, no. 9, 2020, doi: http://dx.doi.org/10.14569/IJACSA.2020.0110935

[13] P. Houngue, R. Sagbo, G. Dahoue, and J. Komaclo, "Smart Monitoring System using Internet of Things: Application for Agricultural Management in Benin" International Journal of Advanced Computer Science and Applications(IJACSA), vol. 13, no. 12, 2022, doi: http://dx.doi.org/10.14569/IJACSA.2022.01312114

[14] S.C. Tun, T. Onizuka, P. Tin, M. Aikawa, I. Kobayashi, T.T. Zin, "Revolutionizing Cow Welfare Monitoring: A Novel Top-View Perspective with Depth Camera-Based Lameness Classification," J. Imaging, vol. 10, no. 67, 2024, doi: https://doi.org/10.3390/jimaging10030067

[15] W. Hao, C. Ren, M. Han, L. Zhang, F. Li, Z. Liu, "Cattle Body Detection Based on YOLOv5-EMA for Precision Livestock Farming," Animals, vol. 13, no. 3535, 2023, doi: https://doi.org/10.3390/ani13223535

[16] Ł. Popek, R.. Perz, G. Galiński, "Comparison of Different Methods of Animal Detection and Recognition on Thermal Camera Images," Electronics, vol. 12, no. 270, 2023, doi: https://doi.org/10.3390/electronics12020270

[17] S. Binta Islam, D. Valles, T.J. Hibbitts, W.A. Ryberg, D.K. Walkup, M.R.J. Forstner, "Animal Species Recognition with Deep Convolutional Neural Networks from Ecological Camera Trap Images," Animals, vol. 13, no. 1526, 2023, doi: https://doi.org/10.3390/ani13091526

[18] M. Tan, W. Chao, J.-K. Cheng, M. Zhou, Y. Ma, X. Jiang, J. Ge, L. Yu, L. Feng, "Animal Detection and Classification from Camera Trap Images Using Different Mainstream Object Detection Architectures," Animals, vol. 12, no. 1976, 2022, doi: https://doi.org/10.3390/ani12151976

[19] M. Ahmad, S. Abbas, A. Fatima, G.F. Issa, T.M. Ghazal, M.A. Khan, "Deep Transfer Learning-Based Animal Face Identification Model Empowered with Vision-Based Hybrid Approach," Appl. Sci., vol. 13, no. 1178, 2023, doi: https://doi.org/10.3390/app13021178

[20] Z. Wan, F. Tian, C. Zhang, "Sheep Face Recognition Model Based on Deep Learning and Bilinear Feature Fusion," Animals, vol. 13, no. 1957, 2023, doi: https://doi.org/10.3390/ani13121957

[21] D. Yeleshetty, L. Spreeuwers, Y. Li, "3D Face Recognition for Cows," 2020 International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 2020, pp. 1-5.

[22] F. Xu, J. Gao, X. Pan, "Cow face recognition for a small sample based on Siamese DB Capsule Network," Ieee Access, vol. 10, 2022, pp. 63189-63198, doi: https://doi.org/10.1109/ACCESS.2022.3182806

[23] R.W. Bello, A.Z.H. Talib, A.S.A.B. Mohamed, "Deep learning-based architectures for recognition of cow using cow nose image pattern," Gazi University Journal of Science, vol. 33, no. 3, 2020, pp. 831-844, doi: https://doi.org/10.35378/gujs.605631

[24] G. Li, G.E. Erickson, Y. Xiong, "Individual Beef Cattle Identification Using Muzzle Images and Deep Learning Techniques," Animals, vol. 12, no. 1453, 2022, doi: https://doi.org/10.3390/ani12111453

[25] M. Ramesh, A.R. Reibman, J.P. Boerman, "Eidetic recognition of cattle using keypoint alignment," Electronic Imaging, vol. 279, 2023, pp. 279-1 - 279-6, doi: https://doi.org/10.2352/EI.2023.35.7.IMAGE-279

[26] R.W. Bellom, A.Z. Talib, A.S. Mohamed, D.A. Olubummo, F.N. Otobo, "Image-based individual cow recognition using body patterns," International Journal of Advanced Computer Science and Applications(IJACSA), vol. 11, no. 3, 2020, doi: http://dx.doi.org/10.14569/IJACSA.2020.0110311

[27] B. Wang, X. Li, X. An, W. Duan, Y. Wang, D. Wang, J. Qi, "Open-Set Recognition of Individual Cows Based on Spatial Feature Transformation and Metric Learning," Animals, vol. 14, no. 1175, 2024, doi: https://doi.org/10.3390/ani14081175

[28] A. Fuentes, S. Han, M.F. Nasir, J. Park, S. Yoon, D.S. Park, "Multiview Monitoring of Individual Cattle Behavior Based on Action Recognition in Closed Barns Using Deep Learning," Animals, vol. 13, 2020, doi: https://doi.org/10.3390/ani13122020

[29] L. Fu, S. Li, S. Kong, R. Ni, H. Pang, Y. Sun, T. Hu, Y. Mu, Y. Guo, H. Gong, "Lightweight individual cow identification based on Ghost combined with attention mechanism," Plos one, vol. 17, no. 10, 2022, p.e0275435, doi: https://doi.org/10.1371/journal.pone.0275435

[30] F. Lv, C. Zhang, C. Lv, "Image recognition of individual cow based on SIFT in Lαβ color space," MATEC Web of Conferences, vol. 176, no. 01023, 2018, doi: https://doi.org/10.1051/matecconf/201817601023

[31] H. Hou, W. Shi, J. Guo, Z. Zhang, W. Shen, S. Kou, "Cow Rump Identification Based on Lightweight Convolutional Neural Networks," Information, vol. 12, no. 361, 2021, doi: https://doi.org/10.3390/info12090361

[32] R.W. Bello, A.Z. Talib, A.S.A. Mohamed, "Real-time cow detection and identification using enhanced particle filter," IOP Conference Series: Materials Science and Engineering, vol. 1051, no. 1, 2021, doi: https://doi.org/10.1088/1757-899X/1051/1/012001

[33] F. Okura, S. Ikuma, Y. Makihara, D. Muramatsu, K. Nakada, Y. Yagi, "RGB-D video-based individual identification of dairy cows using gait and texture analyses," Computers and Electronics in Agriculture, vol. 165, no. 104944, 2019, doi: https://doi.org/10.1016/j.compag.2019.104944

[34] A. Pretto, G. Savio, F. Gottardo, F. Uccheddu, G. Concheri, "A novel low-cost visual ear tag based identification system for precision beef cattle livestock farming," Information Processing in Agriculture, vol. 11, no. 1, 2024, doi: https://doi.org/10.1016/j.inpa.2022.10.003

[35] J. Redmon, A. Farhadi, "YOLOv3: An incremental improvement," arXiv:1804.02767, 2018, [online] Available: https://arxiv.org/abs/1804.02767

[36] W. Shen, H. Hu, B. Dai, et al., "Individual identification of dairy cows based on convolutional neural networks," Multimed Tools Appl, vol. 79, 2020, pp. 14711-14724, doi: https://doi.org/10.1007/s11042-019-7344-7

[37] J. Wanget al., "Open Pose Mask R-CNN Network for Individual Cattle Recognition," IEEE Access, vol. 11, 2023, pp. 113752-113768, doi: https://doi.org/10.1109/ACCESS.2023.3321152

A Modified Lightweight DeepSORT Variant for Vehicle Tracking

Ayoub El-alami[1], Younes Nadir[2], Khalifa Mansouri[3]
Signaux, systèmes distribués et Intelligence Artificielle (M2S2I), Equipe Systèmes Informatiques Distribués (SID), Ecole Normale Supérieure de l'Enseignement Technique (ENSET), Université Hassan II de Casablanca (UH2C), Mohammedia, Morocco[1, 3]
Signaux, systèmes distribués et Intelligence Artificielle (M2S2I), Équipe Technologies de l'Information et Intelligence Artificielle (T2IA), Ecole Nationale Supérieure de l'Art et de Design (ENSAD), Université Hassan II de Casablanca (UH2C), Mohammedia, Morocco[2]

***Abstract*—Object tracking plays a pivotal role in Intelligent Transportation Systems (ITS), enabling applications such as traffic monitoring, congestion management, and enhancing road safety in urban environments. However, existing object tracking algorithms like DeepSORT are computationally intensive, which hinders their deployment on resource-constrained edge devices essential for distributed ITS solutions. Urban mobility challenges necessitate efficient and accurate vehicle tracking to ensure smooth traffic flow and reduce accidents. In this paper, we present a modified lightweight variant of the DeepSORT algorithm tailored for vehicle tracking in traffic surveillance systems. By leveraging multi-dimensional features extracted directly from YOLOv5 detections, our approach eliminates the need for an additional convolutional neural network (CNN) descriptor and reduces computational overhead. Experiments on real-world traffic surveillance data demonstrate that our method reduces tracking time to 25.29% of that required by DeepSORT, with only a minimal increase over the simpler SORT algorithm. Additionally, it maintains low error rates between 0.43% and 1.69% in challenging urban scenarios. Our lightweight solution facilitates efficient and accurate vehicle tracking on edge devices, contributing to more effective ITS deployments and improved road safety.**

Keywords*—Distributed systems; intelligent transportation systems; edge computing; object tracking

I. Introduction

In the era of smart cities, integrating distributed Intelligent Transportation Systems (ITS) with edge computing marks a significant advancement in urban mobility. By processing data closer to the source, using embedded devices with traffic cameras rather than relying solely on centralized servers, ITS can enable real-time decision-making and responsiveness, even in resource-constrained environments. This decentralized approach improves system scalability, reduces latency, and enhances data privacy by processing sensitive information locally. Multiple Object Tracking (MOT), a core component of ITS, refers to tracking the trajectories of multiple objects across video frames. Recent advances in object detection and deep learning have made tracking-by-detection the dominant approach in MOT [1]. This approach formulates MOT as a data association task, linking newly detected objects with those already being tracked.

Recent Multiple Object Tracking (MOT) systems are generally composed of three key components: the object detector, the embedding model, and the data association algorithm. The object detector localizes and identifies objects of interest within each frame, commonly using deep learning models such as YOLO [1], and Faster R-CNN [2]. The embedding model extracts representative features from the detected objects to capture their appearance characteristics. These embeddings combined with spatiotemporal parameters, are used to associate objects across frames and maintain consistent track identities. Finally, the data association algorithm links objects based on their appearance and spatiotemporal similarities. This is typically achieved through methods such as the Hungarian algorithm, as used in DeepSORT [3], or via deep learning-based techniques like Siamese networks or recurrent neural networks [4].

Current object detection and tracking algorithms, notably YOLOv5 and DeepSORT, have achieved high accuracy in multi-object tracking tasks. However, their computational intensity poses significant challenges for deployment on resource-constrained edge devices essential for distributed ITS. This limitation creates a challenge for the practical application of ITS, where real-time processing is crucial for urban mobility and road safety enhancements. To bridge this gap, we propose a lightweight variant of the DeepSORT algorithm that reduces computational overhead while maintaining tracking accuracy, making it suitable for edge computing environments [5], [6], [7], [8], [9], [10].

In this work, we propose a lightweight tracking algorithm that follows a similar workflow to DeepSORT, with a key modification in the appearance embedding step. Traditionally, this step relies on a dedicated CNN model as a descriptor to calculate appearance similarity between objects. While effective, CNN descriptors are computationally intensive, particularly for resource-constrained devices. To enhance the computational efficiency of the tracking process, we eliminate the CNN descriptor and introduce an alternative strategy for object appearance extraction and association.

Our strategy replaces the appearance features provided by the embedding model with data directly derived from the feature map generated during the initial detection phase. These feature maps, produced by the intermediate layers of the detector,

contain high-level abstract representations of the objects. Experiments demonstrate that these representations are a viable substitute for CNN descriptor features when calculating appearance similarity, reducing computational overhead without sacrificing tracking accuracy.

The rest of this paper is structured as following. Section II discusses related work in object detection and multi-object tracking. In Section III, we present our proposed tracking approach, starting with its background methods (i.e. YOLOv5 and DeepSORT), followed by a detailed explanation of our lightweight DeepSORT variant, including the modified detection and appearance similarity processes. Section IV describes our experimental setup and shows the results. In conclusion section, we summarizing our contributions and suggesting future research directions.

II. Related Work

In this section, we present a review of key works related to object detection and multi-object tracking.

A. Object Detection

Since the introduction of deep convolutional neural networks (CNNs), numerous object detection approaches have emerged, leveraging large open datasets such as COCO [11] and Pascal VOC [12]. Early detectors followed a two-stage process, beginning with region proposal generation, followed by object classification and refinement. Models like Fast R-CNN [13] and Faster R-CNN enhanced both speed and accuracy by improving feature extraction and region proposal methods. Despite their effectiveness, two-stage detectors often face computational inefficiencies due to their sequential processing structure.

The introduction of YOLO (You Only Look Once) [1] in 2016 marked a significant shift towards one-stage detectors, which focus on simultaneous detection and classification. Subsequent versions of YOLO have made considerable strides in performance, particularly in real-time applications, by refining this one-stage detection approach. Recent YOLO versions, such as YOLOv5, YOLOv7-tiny, YOLOv8s, and YOLOv9s, offer lightweight models optimized for devices with limited computational power, making them well-suited for edge devices and real-time object detection tasks [15], [16], [17].

B. Multi-Object Tracking

In recent years, tracking-by-detection (TBD) has emerged as the dominant paradigm in Multiple Object Tracking (MOT), driven by the increasing capabilities of object detectors. In the TBD approach, tracking begins with the detection of target objects in each frame, followed by the extraction of image crops and spatial parameters based on the objects' bounding boxes. Tracking is then achieved by establishing correspondences between objects in consecutive frames using the extracted data. To facilitate this, a similarity matrix is constructed using parameters such as distance, IoU (Intersection over Union), and appearance features. Algorithms like the Hungarian algorithm or greedy algorithm are commonly used to solve this assignment problem.

To further improve tracking performance, state estimation algorithms are employed to predict the positions of tracked objects based on their spatiotemporal characteristics. Notable state estimation methods include Kalman filters, particle filters, and Gaussian processes. Two of the most widely recognized tracking algorithms in MOT are SORT (Simple Online and Realtime Tracking) [18] and DeepSORT.

SORT is an efficient algorithm that uses a Kalman filter for state estimation and Hungarian matching for data association. It focuses solely on objects' positions, making it highly effective in real-time scenarios. However, in more challenging conditions, such as complex motion patterns, occlusions or missed detections, SORT may produce incorrect associations, leading to identity switches and reduced tracking robustness. To address these limitations, DeepSORT extends SORT by incorporating cascade matching for enhanced data association. In addition to spatial similarity, DeepSORT uses appearance similarity to refine associations between detected objects and existing tracks, improving tracking robustness in more complex environments [8], [9], [14], [19].

III. Proposed Tracking Approach

Our proposed approach consists of an optimized tracking algorithm based on DeepSORT, designed to run efficiently in terms of processing time while maintaining high tracking accuracy. Our main contribution is the development of a modified tracking algorithm based on YOLOv5 and DeepSORT, incorporating novel methods for appearance feature extraction and appearance similarity calculation. This innovation allows us to eliminate the need for the CNN model traditionally used in DeepSORT for the appearance-embedding task. By removing the CNN model, we significantly reduce the vehicle tracking time, making the algorithm more suitable for deployment on devices with limited computational resources.

A. Background

1) YOLOv5: In our implementation, we utilized the publicly available YOLOv5s detector, trained on the COCO dataset, with an input image size of 416 × 416. YOLOv5 is a widely-used single-stage object detector, known for its clear and flexible architecture, offering high precision and speed. It is available in five different scales: N, S, M, L, and X, representing Nano, Small, Medium, Large, and Xlarge models, respectively. Each variant maintains the same overall structure, scaling the depth and width to improve detection performance.

A key feature of the YOLO architecture is its grid-based approach, which divides the input image into spatial cells. This division allows YOLOv5 to perform object detection at multiple spatial resolutions simultaneously, enabling the model to detect objects of varying sizes and aspect ratios across the image. Each cell is responsible for predicting the bounding box coordinates and associated class probabilities for objects within its spatial region, as illustrated in Fig. 1. For our experiments, we chose the small model (YOLOv5s) due to its optimal trade-off between precision and speed.

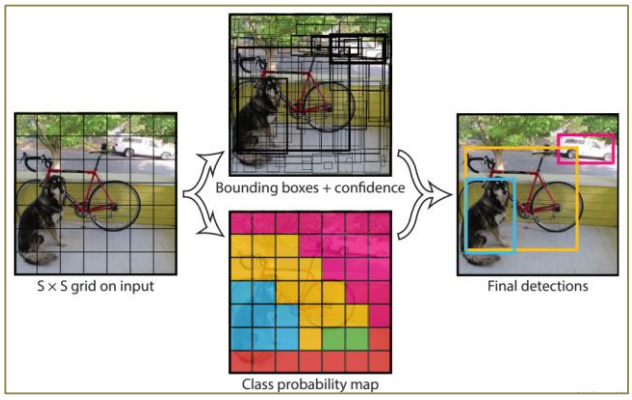


Fig. 1. Example of multiple steps in grid-based approach used in YOLO.

As shown in Fig. 2, the three main components of the YOLOv5 architecture typically include:

- Backbone: The backbone is responsible for extracting features from the input image and is composed of multiple convolutional layers arranged hierarchically. These layers progressively capture higher-level features from the image, enabling the detection of objects of interest by analyzing patterns, edges, and textures at different levels of abstraction.
- Neck: Positioned between the backbone and head, the neck is responsible for additional feature fusion and refinement. Its primary function is to aggregate data extracted at various scales and levels of abstraction by the backbone, ensuring that the model can effectively capture both spatial and contextual information. This step enhances the overall detection performance by combining fine details and broader contextual features before passing them to the detection head.
- Head: The head receives the refined feature maps from the neck and processes them to generate the final detection results. Its layers are designed to predict bounding boxes, objectness scores, and class probabilities, which are used to determine the locations, presence, and categories of objects within the image.

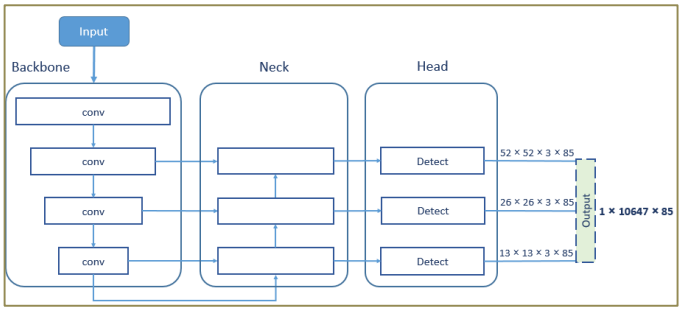


Fig. 2. Abstract overview of YOLOv5 default structure with an input size of 416×416.

2) DeepSORT: DeepSORT builds on the efficiency of the traditional SORT algorithm while incorporating the discriminative power of deep appearance features. Its workflow integrates object detection, feature extraction, data association, and track management to deliver robust tracking results. The DeepSORT workflow can be summarized as follows:

- Detection: The process starts with detecting objects in each frame using an object detector, providing bounding boxes around the detected objects, which serve as input for the tracking process.
- State Estimation: The Kalman Filter (KF) is employed for state prediction, estimating the future states of detected objects based on their current states. This estimation is important for the tracking of each object across frames, especially in cases where an object's detection might be shortly missed.
- Similarity measurement: Multiple metrics are used to calculate object similarity. Mahalanobis distance measures the spatial relationship between predicted states and new detections, while appearance descriptors assess visual similarity. Intersection Over Union (IoU) evaluates the overlap between bounding boxes. These metrics are combined to ensure accurate object association for effective tracking.
- Data Association: The Hungarian algorithm is used to associate detected objects with existing tracks, based on predicted states from the Kalman filter and appearance similarities. This assignment problem combines two metrics: object location and appearance information. DeepSORT's cascade matching prioritizes more frequently seen objects, reducing unstable tracks caused by missed detections. In the final stage, an IoU association is performed for any remaining unmatched targets.
- Track Management: After data association, tracks are updated with their associated detections. New tracks are initialized for unassociated detections, and existing tracks are updated for the next frame using the Kalman filter.

B. Proposed Lightweight Deep SORT Variant

To implement our lightweight variant of DeepSORT, we modified existing modules and introduced new ones to optimize the overall tracking process. The key adjustments made to transition from DeepSORT to our lightweight version include:

- Simultaneous detection and appearance feature extraction: We streamlined the detection process by enabling the simultaneous extraction of both detection data and appearance features. This eliminates the need for a separate embedding model, significantly reducing the computational overhead.
- Adaptation of the appearance similarity process: We adapted the appearance similarity calculation to match the new format of the features obtained from the detection phase. This adjustment ensures seamless integration with our new feature extraction process.

In the following, we provide a detailed explanation of the key adjustments made to our algorithm.

1) Modified detection process: Originally, DeepSORT employs two CNN models: the first functions as an object detector, returning bounding boxes around targets, while the second is a simple CNN descriptor that generates appearance features for those bounding boxes. As mentioned earlier, our approach eliminates the need for a separate CNN model. To maintain appearance association in our algorithm without using a CNN descriptor, we replace the appearance features generated by this descriptor with feature map data associated with the bounding boxes. These feature maps are extracted by YOLOv5's backbone during the detection phase. To extract both feature map data and detection data simultaneously, we implemented several key steps (as shown in Fig. 3), including:

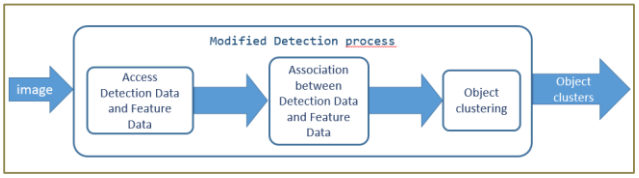


Fig. 3. Workflow of our proposed tracking system.

a) Access detection data and feature data from YOLO: This step involves modifying the model inference process to access intermediate layers of the YOLOv5 architecture, which contain feature map data, in addition to the final detection layer (as shown in Fig. 4). This modification produces an output that includes both detection data and feature map data.

- Detection Data: The detection output includes bounding boxes and class probabilities for small, medium, and large objects, aggregated into a tensor of shape (1, 10647, 85). Specifically, the dimensions 52×52×3×85, 28×28×3×85, and 13×13×3×85 correspond to the detection outputs for small, medium, and large objects, respectively. Here, the grid cell size (e.g., 52×52) refers to the resolution of the detection, the number 3 represents the number of anchor boxes per grid cell, and the 85 elements represent the prediction output for each detection.
- Feature Map Data: The feature map output consists of three tensors representing the features extracted from the backbone layers for each object size. These feature maps are organized in grids of vectors, where each vector is a feature vector for a grid cell. The backbone generates three different vector sizes (128, 256, and 512) for small, medium, and large objects, respectively.

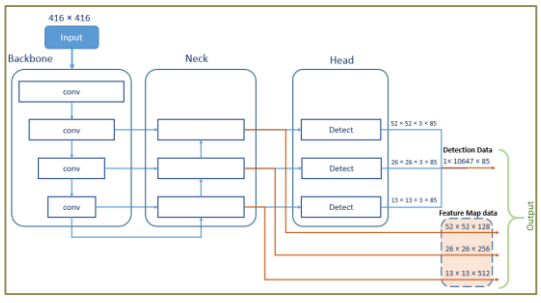


Fig. 4. A representation of the output data extracted from YOLOv5 layers in our algorithm.

b) Association between detection data and feature data: After the model makes its predictions, an association step is performed to map each detected object to its corresponding grid cell in the feature maps to identify its appearance feature. This process involves identifying the spatial locations of object centers within the grid. Since YOLOv5 uses different grid sizes for detection (small (26x26), medium (52x52), and large (13x13)), each object may be represented by multiple detection instances across different grids. Each detection instance is associated with a vector of dimension 85, which includes the object's bounding box coordinates and class information. The feature map data from the grid cell that contains the center of the bounding box is used to represent the object (Fig. 5). Subsequently, the grid cells are mapped to bounding boxes where the box center lies within the grid cell region. This step ensures that all relevant backbone feature vectors are accessible for each detected object.

c) Object clustering: Typically, a non-maximum suppression (NMS) process is applied to remove redundant bounding boxes and retain only the most confident detection. However, in our approach, we aim to extract both bounding box and feature map data. To achieve this, we replace NMS with an Object Clustering (OC) module. The OC module works similarly to NMS, using the IoU (Intersection over Union) metric to compare object overlap. However, instead of eliminating redundant bounding boxes, the OC clusters detections of the same object. This approach allows us to retain all associated feature data, which contributes more effectively to the appearance-based association process.

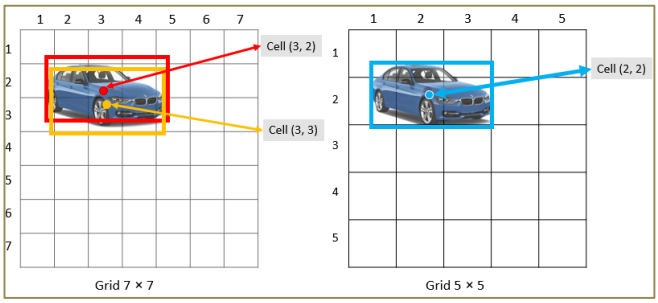


Fig. 5. Example of cell location on different feature grids.

2) Modified appearance similarity process: In the original DeepSORT, a feature bank (or gallery) mechanism is employed to retain long-term appearance information about targets, aiding in recovering object identities after long-term occlusions. This mechanism calculates the appearance similarity between a detection and a tracked target by storing up to 100 appearance descriptors for each track. The similarity is determined by the smallest cosine distance between the detection's appearance vector and the stored vectors in the feature bank. These appearance descriptors are obtained through a pre-trained CNN embedding model. However, this CNN model in DeepSORT contains 2,800,864 parameters and processing a batch of 32 bounding boxes takes approximately 30 ms on an Nvidia GeForce GTX 1050 mobile GPU. On embedded devices or devices with limited resources, especially those without GPU capability, the processing time may increase significantly.

To optimize the tracker for such devices, we propose extracting the appearance similarity without relying on the CNN model. Instead, we leverage the feature map vectors generated during the detection process. As discussed earlier, our detection process associates each detected object with its corresponding feature map vectors extracted from the YOLOv5 neck. The number of feature vectors varies depending on the number of predicted detections for that object, and these vectors differ in dimensionality based on the three grid sizes used in YOLOv5's detection process. These feature vectors are stored for each object in what we refer to as an object cluster. An object cluster consists of all the feature vectors associated with the multiple predicted detections of the same object resulted from the previous modified detection phase.

For example, in YOLOv5s, the grid for small objects (52x52) generates feature vectors with a dimensionality of 128, the grid for medium-sized objects (26x26) produces vectors with a dimensionality of 256, and the grid for large objects (13x13) creates vectors with a dimensionality of 512. Considering these feature vectors, we propose an approach that evaluates appearance similarity by comparing vectors of the same size between two objects.

Algorithm 1 outlines the steps for calculating appearance similarity between a detection D and a track T. First, we divide the feature vectors into three sets based on their dimensions (128, 256, and 512). Then, we compute the similarity score for each set as the minimum cosine distance between corresponding vectors. The final appearance similarity score is determined by taking the overall minimum of these three scores.

$$Simil(DF_{Dim}, TF_{Dim}) = \min\{\cos(df_i, tf_j) \mid df_i \in DF_{Dim}, tf_j \in TF_{Dim}\} \quad \text{(Eq.1)}$$

Algorithm 1 Appearance Similarity between two objects

Input: $DF = \{df_1, \ldots, df_n\}$, // features of object 1
$TF = \{tf_1, \ldots, tf_n\}$ // features of object 2

1: $(DF_{128}, DF_{256}, DF_{512})$ = *separate_by_size*(DF)
2: $(TF_{128}, TF_{256}, TF_{512})$ = *separate_by_size*(TF)
3: $simil_{128}$ = Simil(DF_{128} , TF_{128}) using Eq. 1
4: $simil_{256}$ = Simil(DF_{256} , TF_{256}) using Eq. 1
5: $simil_{512}$ = Simil(DF_{512} , TF_{512}) using Eq. 1
6: **return** min{ $simil_{128}$, $simil_{256}$, $simil_{512}$ }

IV. EXPERIMENTS

To evaluate the performance of our proposed tracking algorithm, we conducted a series of experiments using real-world highway surveillance videos. We compared the tracking accuracy, robustness, and computational efficiency of our algorithm against the state-of-the-art SORT and DeepSORT algorithms. The video sequences used for testing captured a variety of challenging traffic scenarios, including high-density traffic and occlusions, to thoroughly assess the algorithm's performance under realistic conditions.

For our experiments, we utilized a locally collected traffic video dataset comprising over 10,000 frames. This dataset captures real-world urban traffic scenarios with varying numbers of vehicles, ranging from light traffic to highly congested conditions. It includes challenging situations such as shadowing caused by varying lighting conditions throughout the day and occlusions resulting from vehicles overlapping or obstructing each other. The videos were recorded at a resolution of 1920x1080 and a frame rate of 30 fps, ensuring high-quality imagery for accurate detection and tracking. The dataset encompasses different times of the day and weather conditions, providing a comprehensive set of challenges for evaluating the robustness and scalability of our proposed tracking approach.

The experiments were performed on a Windows machine equipped with an Intel Xeon Silver 4110 CPU @ 2.10 GHz and 32 GB of RAM. To simulate low-resource environments, we deployed our tracker within a Docker container configured with reduced resources, specifically 4 CPU cores and 4 GB of RAM, to mimic the capabilities of edge devices. This setup allowed us to assess the feasibility of deploying our algorithm on resource-constrained platforms without compromising its real-time performance.

To evaluate the computational efficiency of our algorithm, we measured the average processing time required to track varying numbers of vehicles in the scene. This processing time represents the duration needed to perform the tracking task for each frame, excluding the object detection phase. This metric provided valuable insights into the performance of our algorithm, particularly in terms of calculating appearance similarity and data association, under different workload conditions.

To assess the scalability of our method, we analyzed its performance across scenes with varying numbers of vehicles. Table I illustrates the processing time per frame (in milliseconds) relative to the number of detected vehicles compared to DeepSORT and SORT. Additionally, Fig. 6 provides a graphical representation of the data in Table I, offering a visual comparison of the processing times between the algorithms. Our findings revealed that, on average, the processing time of our tracker is about 25 % of the time required by DeepSORT, and approximately 10% higher than SORT. These results underscore the significant computational efficiency of our tracker, as it maintains near-constant processing time as the number of vehicles increases, demonstrating efficient scalability. This consistent performance is critical for real-time applications in ITS, where traffic density can fluctuate significantly.

TABLE I. MEASURED PROCESSING TIMES IN MILLISECONDS FOR EACH ALGORITHM WITH VARYING NUMBERS OF VEHICLES

# Number of Vehicles	DEEP SORT	SORT	Our Tracker
1	24,58	2,23	**2,97**
2	39,75	3,82	**7,01**
3	52,43	5,80	**12,77**
4	65,14	7,85	**20,71**
5	73,83	10,37	**31,21**
6	87,07	12,38	**38,08**
7	96,49	12,69	**42,06**
Avg Proce. time	**45,01**	**6,83**	**11,38**

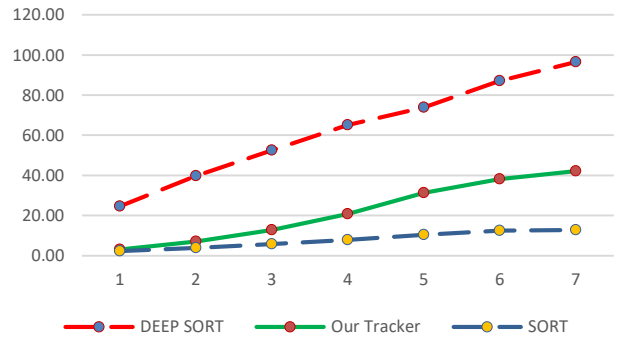


Fig. 6. Visual comparison of the processing times between our algorithm and the DeepSORT and SORT.

To evaluate the accuracy of our tracker, instead of using manually annotated video sequences with ground truth object trajectories, we opted to compare our tracking results directly with those of DeepSORT, which achieved a perfect score in our scenarios. We employed an error rate metric, calculated by summing the number of misses and false matches, then dividing by the total number of matches. This metric provides a comprehensive assessment of the accuracy and reliability of our algorithm under various conditions, with lower error rates indicating higher tracking accuracy.

To assess the robustness of our tracker, we introduced real-world uncertainty by simulating missed detections with a certain probability. We utilized a parameter called the skip detection rate, which we tested with three values: 0%, 25%, and 50%. This parameter simulates scenarios of missed detections by occasionally sending an empty detection output to the tracker. This test allowed us to evaluate the tracker's ability to re-identify objects through appearance association, even after missed detections, providing valuable insights into its reliability for real-world applications.

TABLE II. A COMPARISON BETWEEN THE ERROR RATES OF OUR TRACKER AND SORT REGARDING VARIOUS SKIP RATE VALUES

Skip rate	**0%**		**25%**		**50%**	
Algorithm	Our tracker	SORT	Our tracker	SORT	Our tracker	SORT
Miss matches	34	92	53	1048	50	1164
False matches	2	0	6	2	14	0
Total matches	8369	8362	6119	5949	3783	3661
Error Rate	**0,43%**	1,10 %	**0,96 %**	17,65 %	**1,69 %**	31,79 %

Table II presents the accuracy performance results of our tracker compared to SORT across three scenarios of skip detection rates: 0%, 25%, and 50%. DeepSORT was excluded from the comparison due to its perfect accuracy score in all scenarios.

The results show that under normal conditions (i.e., without skip detections) both our tracker and SORT achieve high accuracy with very low error rates. Notably, our tracker achieves an error rate of 0.43%, outperforming SORT, which records an error rate of 1.10%. In more challenging scenarios with non-zero skip detection rates, our tracker demonstrates a significant advantage. At skip rates of 25% and 50%, our tracker achieves error rates of 0.96% and 1.69%, respectively, while SORT shows much higher error rates of 17.65% and 31.79%. These results highlight the superior robustness of our tracker in handling missed detections compared to SORT (Fig. 7).

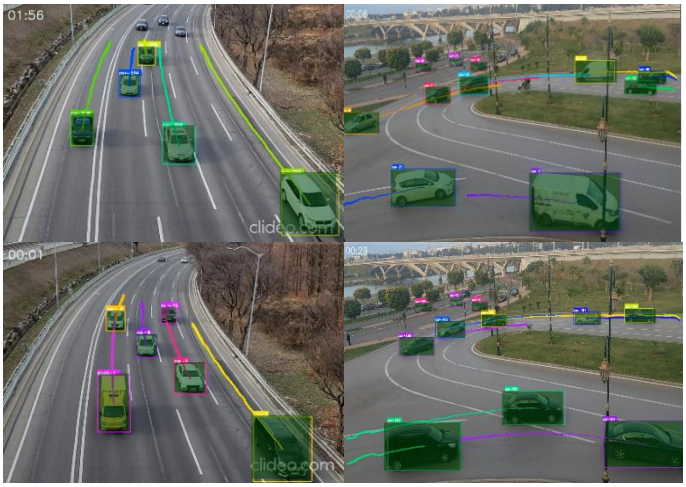


Fig. 7. Sample of tracking results visualization using our lightweight tracker.

V. DISCUSSION

The experimental results confirm that our modified lightweight DeepSORT variant significantly reduces computational requirements while maintaining high tracking accuracy. This improvement addresses a critical barrier in deploying effective ITS solutions on edge devices, enabling more widespread and efficient traffic monitoring systems. Our approach demonstrates that leveraging features directly from YOLOv5 detections is an effective strategy for reducing overhead without sacrificing performance. This finding suggests a paradigm shift in multi-object tracking, where the integration of detection and tracking components can lead to more streamlined and efficient algorithms.

One of the key insights from our study is that integrating detection and appearance features can lead to more efficient multi-object tracking solutions suitable for real-time deployment. However, we acknowledge that testing on a single, localized dataset may limit the generalizability of our findings. Evaluating the method on additional datasets, possibly from different regions or with different characteristics, would further substantiate its scalability.

One challenge we observed is that in highly congested scenes, the absence of an additional CNN descriptor slightly affected the appearance similarity process, occasionally leading to some miss matching and ID switches. This issue could be addressed in future work by incorporating further improvements to the processing method. Some of these improvements could include exploring optimized adaptive association mechanisms and integrating more recent and enhanced detection models to further improve performance. We also we aims to investigate the integration of our tracking algorithm with other ITS components, such as traffic prediction models and anomaly detection systems.

A promising direction for improvement is training the detector on a specialized dataset enriched with detailed annotations. By labeling objects with attributes like distinct colors, shapes, and varied viewing angles, the detector's backbone network can learn to recognize more discriminative features. This enhancement would significantly improves its ability to differentiate between objects, particularly in complex traffic scenarios with similar-looking vehicles or difficult perspectives, leading to more accurate and reliable tracking.

VI. CONCLUSION

In this paper, we present a novel lightweight tracking algorithm designed specifically for vehicle tracking in Intelligent Transportation Systems (ITS). By leveraging the predictability of vehicle trajectories and optimizing the DeepSORT workflow, we propose an algorithm that strikes a balance between accuracy and computational efficiency, making it well-suited for deployment on resource-constrained edge devices. Our approach utilizes features extracted from the detector network as a source of appearance information for the appearance matching task, effectively eliminating the need for the traditional CNN descriptor used in DeepSORT. This significantly reduces the computational load of our tracking algorithm. Experiments with real-world traffic surveillance data reveal that our tracker not only outperforms traditional methods like SORT in tracking accuracy but also significantly reduces processing time compared to DeepSORT. This efficiency is crucial for real-time applications on resource-constrained devices. However, some scenarios when a large number of features are extracted with objects, the algorithm can encounter some challenges, as the similarity calculation becomes computationally demanding. To address this, we plan in future works to add a new feature limitation mechanism that adjust the number of features used in the similarity processing, allowing users to tradeoff between efficiency and accuracy. Additionally, we aim to enhance the detector by training it on specialized datasets to improve object distinction. We also intend to evaluate the effectiveness of our method on devices like the Raspberry Pi to assess its performance on typical edge computing hardware and explore its applicability in more general use cases within ITS.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, et A. Farhadi, « You Only Look Once: Unified, Real-Time Object Detection », in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA: IEEE, juin 2016, p. 779-788. doi: 10.1109/CVPR.2016.91.

[2] S. Ren, K. He, R. Girshick, et J. Sun, « Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks », in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2015. Consulté le: 19 septembre 2022. [En ligne]. Disponible sur: https://proceedings.neurips.cc/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html

[3] N. Wojke, A. Bewley, et D. Paulus, « Simple Online and Realtime Tracking with a Deep Association Metric », 21 mars 2017, arXiv: arXiv:1703.07402. doi: 10.48550/arXiv.1703.07402.

[4] Z. Zhang et H. Peng, « Deeper and Wider Siamese Networks for Real-Time Visual Tracking », présenté à Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, p. 4591-4600. Consulté le: 28 avril 2024. [En ligne]. Disponible sur: https://openaccess.thecvf.com/content_CVPR_2019/html/Zhang_Deeper_and_Wider_Siamese_Networks_for_Real-Time_Visual_Tracking_CVPR_2019_paper.html

[5] C. Duan et X. Li, « Multi-target Tracking Based on Deep Sort in Traffic Scene », J. Phys.: Conf. Ser., vol. 1952, no 2, p. 022074, juin 2021, doi: 10.1088/1742-6596/1952/2/022074.

[6] H. A. Abdelali, H. Derrouz, Y. Zennayi, R. O. H. Thami, et F. Bourzeix, « Multiple Hypothesis Detection and Tracking Using Deep Learning for Video Traffic Surveillance », IEEE Access, vol. 9, p. 164282-164291, 2021, doi: 10.1109/ACCESS.2021.3133529.

[7] M. Anandhalli, V. P. Baligar, P. Baligar, P. Deepsir, et M. Iti, « Vehicle detection and tracking for traffic management », IJ-AI, vol. 10, no 1, p. 66, mars 2021, doi: 10.11591/ijai.v10.i1.pp66-73.

[8] A. El-Alami, Y. Nadir, L. Amhaimar, et K. Mansouri, « An efficient hybrid approach for vehicle detection and tracking », in 2023 10th International Conference on Wireless Networks and Mobile Communications (WINCOM), oct. 2023, p. 1-8. doi: 10.1109/WINCOM59760.2023.10322924.

[9] A. El-Alami, Y. Nadir, et K. Mansouri, « A hybrid vehicle tracking System for Low-power Embedded Devices », in 2024 International Conference on Circuit, Systems and Communication (ICCSC), juin 2024, p. 1-6. doi: 10.1109/ICCSC62074.2024.10617125.

[10] Z. Charouh, A. Ezzouhri, M. Ghogho, et Z. Guennoun, « A Resource-Efficient CNN-Based Method for Moving Vehicle Detection », Sensors, vol. 22, no 3, Art. no 3, janv. 2022, doi: 10.3390/s22031193.

[11] T.-Y. Lin et al., « Microsoft COCO: Common Objects in Context », 20 février 2015, arXiv: arXiv:1405.0312. Consulté le: 18 juin 2023. [En ligne]. Disponible sur: http://arxiv.org/abs/1405.0312

[12] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, et A. Zisserman, « The Pascal Visual Object Classes (VOC) Challenge », Int J Comput Vis, vol. 88, no 2, p. 303-338, juin 2010, doi: 10.1007/s11263-009-0275-4.

[13] R. Girshick, « Fast R-CNN », 27 septembre 2015, arXiv: arXiv:1504.08083. Consulté le: 19 septembre 2022. [En ligne]. Disponible sur: http://arxiv.org/abs/1504.08083

[14] D. V. Tu, P. M. Quang, H. P. Nghi, et T. N. Thinh, « An Edge AI-Based Vehicle Tracking Solution for Smart Parking Systems », in Intelligence of Things: Technologies and Applications, N.-N. Dao, T. N. Thinh, et N. T. Nguyen, Éd., in Lecture Notes on Data Engineering and Communications Technologies. Cham: Springer Nature Switzerland, 2023, p. 234-243. doi: 10.1007/978-3-031-46573-4_22.

[15] A. Benjumea, I. Teeti, F. Cuzzolin, et A. Bradley, « YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles », 3 janvier 2023, arXiv: arXiv:2112.11798. Consulté le: 7 avril 2024. [En ligne]. Disponible sur: http://arxiv.org/abs/2112.11798

[16] A. El-Alami, Y. Nadir, et K. Mansouri, « A review of object detection approaches for traffic surveillance systems », International Journal of Electrical and Computer Engineering (IJECE), vol. 14, no 5, Art. no 5, oct. 2024, doi: 10.11591/ijece.v14i5.pp5221-5233.

[17] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, et B. Lee, « A Survey of Modern Deep Learning based Object Detection Models », 12 mai 2021, arXiv: arXiv:2104.11892. Consulté le: 7 septembre 2022. [En ligne]. Disponible sur: http://arxiv.org/abs/2104.11892

[18] A. Bewley, Z. Ge, L. Ott, F. Ramos, et B. Upcroft, « Simple Online and Realtime Tracking », in 2016 IEEE International Conference on Image Processing (ICIP), sept. 2016, p. 3464-3468. doi: 10.1109/ICIP.2016.7533003.

[19] M. Elhoseny, « Multi-object Detection and Tracking (MODT) Machine Learning Model for Real-Time Video Surveillance Systems », Circuits Syst Signal Process, vol. 39, no 2, p. 611-630, févr. 2020, doi: 10.1007/s00034-019-01234-7.

Multi-Site Cross Calibration on the LAPAN-A3/IPB Satellite Multispectral Camera with One-Dimensional Kalman Filter Optimization

Sartika Salaswati[1], Adhi Harmoko Saputro[2], Wahyudi Hasbi[3], Deddy El Amin[4], Patria Rachman Hakim[5], Silmie Vidiya Fani[6], Agung Wahyudiono[7], Ega Asti Anggari[8]
Department of Physics, Faculty of Mathematics and Natural Sciences, University of Indonesia, Indonesia[1, 2]
Research Center for Satellite Technology, National Research and Innovation Agency (BRIN), Indonesia[1, 3, 4, 5, 7, 8]
Research Center for Space, National Research and Innovation Agency (BRIN), Indonesia[6]

***Abstract*—Multispectral cameras on remote sensing satellites must have good radiometric quality due to their wide range of applications. One type of radiometric calibration that can be performed while the satellite is in orbit is cross-calibration. This research focuses on cross-calibration because it has advantages, including being cost-effective and capable of frequent execution. We proposed a multi-site cross-calibration method with two reference cameras using six calibration sites in 2023. The LISA LAPAN-A3 (LA3) camera serves as the target camera, while the OLI LANDSAT-8 (OL8) and MSI SENTINEL-2 (MS2) cameras act as the reference cameras. The calibration process results in numerous calibration coefficients for each channel, thus requiring optimization to produce a single calibration coefficient. The optimization process uses a one-dimensional Kalman filter to reduce measurement noise. The results show that the one-dimensional Kalman filter can reduce noise in the calibration coefficient data, making LA3 radiance values closer to the reference radiance values. Additionally, this study demonstrates that LA3 calibration results with MS2 as the reference camera are better than those with OL8 as the reference.**

Keywords*—Cross-calibration; multi sites; multispectral camera; LAPAN-A3/IPB Satellite; LISA LAPAN-A3 (LA3); OLI LANDSAT-8 (OL8); MSI/SENTINEL-2 (MS2); one-dimensional Kalman filter

I. INTRODUCTION

Line Imager Space Application (LISA) is a multispectral camera on the LAPAN A3/IPB satellite. This camera is the main payload that supports the remote sensing mission of the LAPAN-A3/IPB satellite. This camera has blue, green, red, and near-infrared (NIR) channels, with a ground resolution of 15 meters and 120 km swath-width from 510 km altitude. Besides, this camera has 16-bit radiometric resolution and 21 days of temporal resolution [1].

Multispectral cameras on satellites are extensively utilized in remote sensing applications, encompassing the detection and characterization of agroforestry systems [2], evaluation of land-use changes [3], identification of terrestrial vegetation [4], mapping of mangrove forests [5], assessment of forest restoration [6], bamboo mapping [7], and evaluation for sustainable urban development [8].

LISA LAPAN-A3 (LA3), similar to the multispectral cameras used in other remote sensing satellites, serves multiple purposes, including the identification of plant types and growth conditions [9], monitoring drought areas [10], calculating chlorophyll in plants [11], monitoring rice fields [12], and performing land use analysis [13].

This camera produces images consisting of a collection of pixels with specific values. The values in these pixels are called digital numbers (DN). Digital numbers represent the radiometric values of objects in an image [14]. However, calibration coefficients are needed to convert the digital number (DN) values to obtain TOA radiance or TOA reflectance. An absolute radiometric calibration process is required to generate calibration coefficients for the images.

There are several methods of absolute radiometric calibration, including the on-board calibration method [15][16], the Rayleigh calibration method [17][18][19], the vicarious calibration method [16][20][21][22][23], and the cross-calibration method [16][24][25]. Based on previous literature reviews, the cross-calibration method is considered an efficient alternative as it has the advantage of being low cost, high calibration frequency, and can be carried out repeatedly to obtain calibration coefficients for each channel [26].

Furthermore, research on cross-calibration of satellite cameras has been widely conducted. Lu researched cross-calibration for medium-resolution multispectral cameras with large angles [24]. Gao researched cross-calibration to monitor VNIR sensor degradation on the Gaofen-4/GF-4 satellite using three reference sensors, namely OLI/Landsat-8, MSI/Sentinel 2, and MODIS/Terra [25]. Jie Han researched cross-calibration based on a radiometric block adjustment (RBA) algorithm to minimize radiometric effects due to differences in integration time [27]. Mizuochi researched cross calibration with inter-band calibration on a hyperspectral camera [28]. Kohei Arai researched cross-calibration based on vicarious calibration measurements [29]. Due to the significant developments related to this method, this study will also develop a radiometric calibration method based on cross-calibration.

The LA3 calibration has been the subject of numerous investigations. Laboratory calibration [30], vicarious calibration at the Kupang cement mine [31], and Jaddih Hill Madura [32][33] are some of the studies that fall under this category. Although these investigations have yielded calibration

coefficients, the current calibration methods still have drawbacks.

Additionally, the calibration coefficients for satellites need to be updated because there is a possibility that satellites degrade over time. Therefore, other methods are required to complement previous research. This study proposes a multi-site cross-calibration method for the LA3 camera with one-dimensional Kalman Filter for the optimization process. This method is a cross-calibration technique developed from several previously reviewed references. The proposed method differs from the usual cross-calibration techniques because this method uses more than one calibration site, and the calibration coefficients are optimized using the Kalman Filter, resulting in a single final calibration coefficient.

In this study, cross-calibration was also performed using two types of reference cameras, namely, OLI LANDSAT-8 (OL8) and MSI SENTINEL-2 (MS2). The calibration coefficients from these two references were then validated using a cross-validation method, so the calibration result of LA3 with OL8 reference was validated with MS2, and the calibration result of LA3 with MS2 reference was validated with OL8. This research aims to determine if this method can be successfully implemented on the LAPAN-A3/IPB satellite's multispectral camera, resulting in radiance values that are closely comparable to the standard values. This method is expected to complement previous methods and should be applied to the LA3 camera periodically.

II. Descriptions of Calibration Sites and Datasets

A. Calibration Sites

In this study, six sites were used: five in North Africa and one in France. The sites in North Africa are Libya 1, Libya 4, Algeria 3, Mauritania 1, and Mauritania 2. The site in France is the La Crau calibration site. Libya-1, Libya-4, and Algeria-3 are deserts with dunes 100 meters high and a spatial variability of <3% over 100 km. Mauritania 1 and 2 have lower spatial variability compared to the others. The La Crau site has a spatial variability of <3% and a temporal variability of <2% over an area of <20 km [34].

La Crau is one of the RadCalNet calibration sites that provides free radiometric data. The data available includes TOA reflectance, continuously updated every 30 minutes and at a sampling interval of 10 nm within the 380 nm – 2500 nm spectral range. Each site has automated instruments that can generate reflectance and atmospheric data for that site. The provided data is guaranteed in terms of quality and consistency according to international standards [35]. According to research by Revel et al., this area is more heterogeneous than the Gobabeb site in Namibia, a desert [36]. Table I contains the list of calibration sites used in this study.

TABLE I. Calibration Sites

	Sites	Latitude (°)		Longitude (°)	
		Max	***Min***	***Max***	***Min***
PICS	Algeria-3	30.50	30.20	7.60	7.30
	Libya-1	24.80	24.50	13.60	13.30
	Libya-4	28.65	28.42	23.02	22.80
	Mauritania-1	19.60	19.40	-9.20	-9.40
	Mauritania-2	20.60	20.40	-8.50	-8.80
Rad-Cal-Net	La Crau, France	43.56	43.55	4.86	4.86

B. Datasets

The images used in this study are from the LA3 camera as the target camera, and from the OL8 and MS2 cameras as reference cameras. The images are from the calibration sites mentioned earlier, taken in 2023. There are 8 images with 66 pairs of ROIs used for each reference camera.

LA3 is a multispectral camera on the LAPAN-A3 satellite with a wavelength range of 410 nm – 900 nm. This camera consists of four channels: red, green, blue, and NIR. It has a resolution of 15 m, a swath width of 120 km, and a revisit time of 21 days. Based on these specifications, OL8 and MS2 were selected as reference cameras because their specifications are quite similar to those of the LA3 camera.

OL8 is a multispectral camera on the LANDSAT-8 satellite with a wavelength range of 430 nm – 2290 nm. This camera consists of 8 channels with a resolution of 30 m and 1 channel with a resolution of 15 m. The 8 channels with a resolution of 30 m include coastal aerosol, blue, green, red, NIR, SWIR1, SWIR2, and cirrus. The 15 m resolution channel is the panchromatic channel. The camera has a swath width of 185 km and a revisit time of 16 days [37] [38].

MS2 is a multispectral camera on the SENTINEL-2 satellite with a wavelength range of 442 nm – 2202 nm. This camera consists of four channels with a resolution of 10 m, six channels with a resolution of 20 m, and three channels with a resolution of 60 m. SENTINEL-2 is a satellite that consists of two identical twin spacecraft in the same orbit with a 180° phase difference. This satellite's revisit period of five days is relatively short due to the presence of two twin satellites. The mission of this satellite is a multispectral imaging mission with a swath width of 290 km [39]. Table II displays the specs for the reference and target cameras at RGBN wavelengths.

TABLE II. Specifications for Multispectral Cameras on Satellites

Specifications	**OLI/Landsat 8**	**MSI/Sentinel 2**	**LISA/LAPAN-A3**
Resolution	30 m	10 m	15 m
Swath Width	185 km	290 km	120 km
Spectral			
Red	0.64 – 0.67 µm	0.65 – 0.68 µm	0.63 – 0.70 µm
Green	0.53 – 0.59 µm	0.54 – 0.58 µm	0.51 – 0.58 µm
NIR	0.85 – 0.88 µm	0.78 – 0.90 µm	0.77 – 0.90 µm

Additionally, reflectance data from the spectral library and aerosol optical depth data from the AERONET site are used as supporting data. Reflectance data—specifically, desert reflectance—is taken from the spectral library. The region nearest to the calibration sites is where the aerosol data is derived from. These two data sets serve as inputs for the MODTRAN software to generate the surface reflectance of the calibration sites.

III. CALIBRATION METHOD

The proposed calibration method is a cross-calibration technique with radiometric parameters using several internationally standardized calibration sites (multi-sites). The image acquisition of these sites was conducted throughout 2023. Two reference cameras, OL8 and MS2, are also used in this procedure. Fig. 1 displays the block diagram for the multi-site cross-calibration system for the LA3 camera.

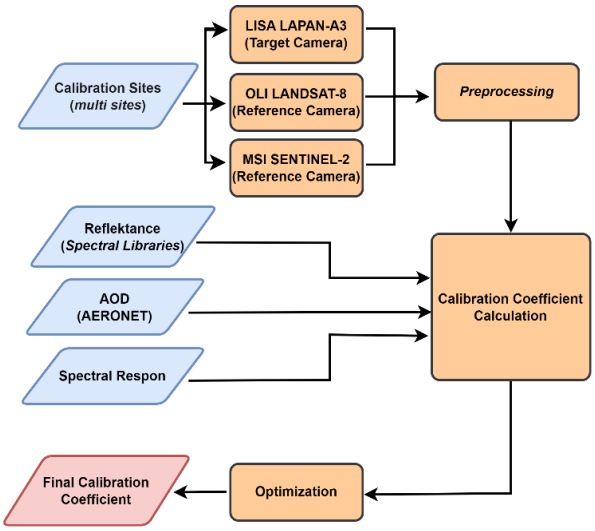


Fig. 1. Block diagram of the multi-site cross-calibration system on the LISA LAPAN-A3 (LA3).

Based on Fig. 1, the first step of this calibration involves LA3 (target camera), OL8, and MS2 (reference cameras) capturing images of the calibration sites at nearly the same time. Next, the cross-calibration method is used for the preprocessing and calibration coefficient calculation steps. This procedure will yield many calibration coefficients. The Kalman filter is used to conduct an optimization process to generate a single final calibration coefficient. The final calibration coefficient is then determined by selecting the coefficient with the smallest error. This process is carried out for each image channel (RGBN).

A. Preprocessing

Before the calibration process, image preprocessing is conducted on both the target and the reference camera images. For the LA3 (target camera) image, as well as the OL8 and MS2 (reference camera) images, the process involves determining the Region of Interest (ROI).

The ROI determination process for LA3 includes setting the ROI boundary coordinates based on the recommendations of the Committee on Earth Observation Satellites (CEOS), identifying ROI points within the ROI boundary, adjusting the ROI points if they do not match the reference cameras, selecting 4 x 4 pixels at each ROI point, and calculating the average Digital Number (DN) of the selected pixels. This average DN will then be processed further.

Meanwhile, for OL8 and MS2, which are the reference cameras, the ROI determination process involves setting the ROI boundary coordinates based on the recommendations of CEOS, identifying ROI points within the ROI boundary, selecting 2 x 2 pixels for OL8 and 6 x 6 pixels for MS2 at each ROI point, and calculating the average DN of the selected pixels to obtain the reference camera DN. The DN values from the target and reference cameras will be used to calculate calibration coefficients.

B. Calculation of Calibration Coefficients

The calculation of calibration coefficients is carried out under the assumption that the response of the LA3 camera is linear. Therefore, the calibration coefficients can be calculated using Eq. (1) [40].

$$\frac{DN_L}{g_L} + o_L = \frac{k.DN_R.c_R.E_L.cos(\theta_L)}{\pi.d^2.cos(\theta_R)} \quad (1)$$

Where DN_L is the DN for LA3, g_L is the calibration gain for LA3, O_L is the calibration offset for LA3, SBAF is the Spectral Band Adjustment Factor (SBAF), DN_R is the digital number of the reference camera, c_R is the calibration gain of the reference camera, E_L is the average solar irradiance for the channel being considered, θ_L is the solar zenith angle for the LA3 camera, d is the average distance between the sun and the Earth, and θ_R is the solar zenith angle for the reference camera. Based on the results of laboratory calibration, the calibration offset for the LA3 camera is 0, so the calibration coefficient (gain) can be obtained using Eq. (2).

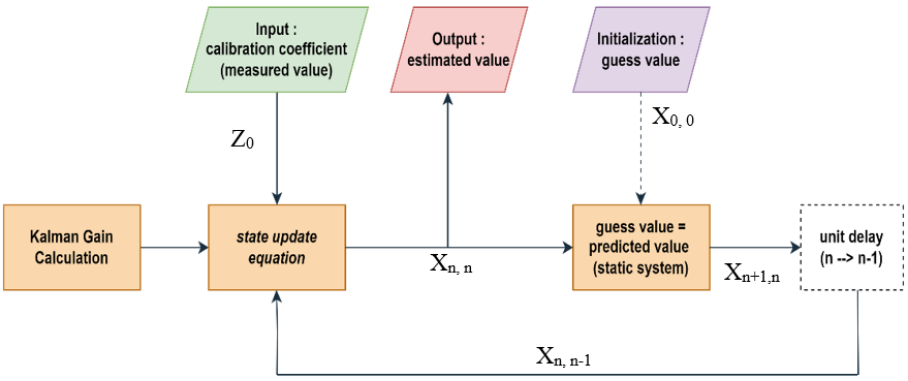


Fig. 2. Block diagram of one-dimensional Kalman filter.

$$g_L = \frac{DN_L . \pi . d^2 . cos(\theta_R)}{SBAF . DN_R . c_R . E_L . cos(\theta_L)} \quad (2)$$

Based on Eq. (2), several inputs and constants are used to obtain the calibration coefficient. DN_L and DN_R are derived from the preprocessing stage. θ_L, θ_R, and c_R come from the image metadata. π, d, and E_L are known constants. Meanwhile, SBAF is a constant obtained from the spectral matching process.

Spectral matching is adjusting spectral responses to account for the differences between the target camera and the reference cameras. This process produces a correction factor known as the Spectral Band Adjustment Factor (SBAF). This correction factor is the ratio of the reflectance of the target camera to that of the reference camera, simulated using MODTRAN software. This process requires inputs such as ground reflectance, aerosol optical depth (AOD), and the spectral response functions of both the target and reference cameras.

The outputs from preprocessing (DNL and DNR), spectral matching (SBAF), and metadata (θL, θR, cR) are then used as inputs for the calculation of the calibration coefficient. The calibration coefficient is calculated based on equation 3.2, resulting in the calibration coefficient (gL). This calculation is carried out for every channel in the LA3 image, yielding a calibration coefficient for each channel (gLR, gLG, gLB, gLN) in the form of a gain factor.

C. Optimization Process

The calibration process will yield many calibration coefficients for each channel. Therefore, an optimization process is needed to obtain the best calibration coefficient. This work uses a one-dimensional Kalman filter to carry out the optimization process. This method is selected because it is anticipated to lower measurement noise, leading to a more accurate radiance measurement by the LA3. Fig. 2 displays the optimization process diagram using the one-dimensional Kalman filter, while Eq. (3) displays the mathematical formula for the output of the one-dimensional Kalman filter.

$$X_{n,n} = X_{n,n-1} + \frac{1}{n}(Z_n - X_{n,n-1}) \quad (3)$$

with

$X_{n,n}$: output (estimated value)

Z_n : measurement value

$X_{0,0}$: predicted value

$X_{n+1,n}$: predicted value

$X_{n,n-1}$: predicted value in the previous process

$\frac{1}{n}$: Kalman gain

The measurement values are the subsequent calibration coefficients obtained from the calibration process, while the initial guess value used is the first calibration coefficient from the measurement process. The calibration coefficient will ultimately be used as the final predicted or estimated value from the optimization process using the one-dimensional Kalman filter.

IV. Result

A. Image Data

This study uses some images with 66 ROI points. The number of ROIs is adjusted based on the image area that can be used for calibration ROI and corresponds to the reference image area. Each image uses 4x4 pixels for the LISA LAPAN-A3 (LA3) image, 2x2 pixels for the OLI LANDSAT-8 (OL8) image, and 6x6 pixels for the MSI SENTINEL-2 (MS2) image, representing a total area of 60x60 meters on the ground. The LA3 images are obtained from the LA3 image database, OL8 images are obtained from the website https://earthexplorer.usgs.gov/, and MS2 images are obtained from the website https://sentinels.copernicus.eu/. The calibration sites and reference image acquisition times are based on the availability of LA3 images. As previously noted, image requirements for calibration are also taken into account. Fig. 3 displays an example of ROI construction on images of the calibration site.

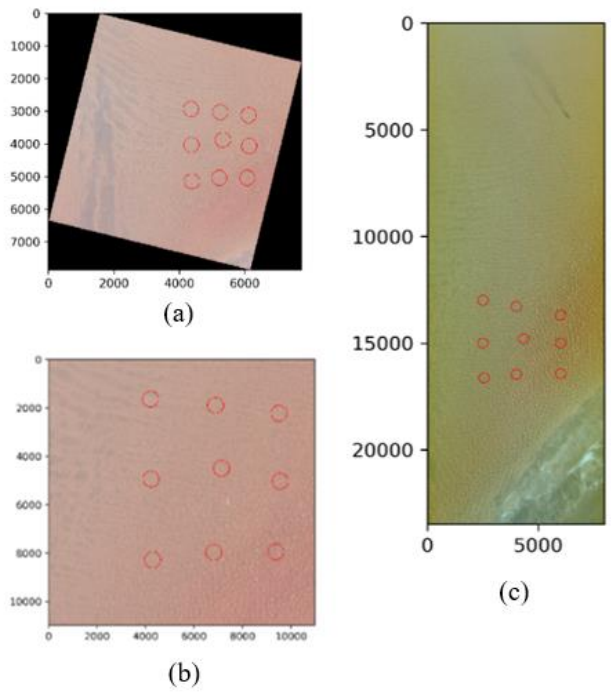


Fig. 3. ROI of Algeria-3 calibration site on imagery citra (a) OL8, (b) MS2, and (c) LA3.

B. Calibration Result

The digital number (DN) for each pixel in each ROI is obtained once the images have been ROI'd. Subsequently, the average DN of the pixels within each ROI is calculated to get a single DN value for each ROI. Every image is subjected to this procedure, including the LA3 image as the target image and OL8 and MS2 as the reference images.

The next step is spectral matching, which yields a spectral band adjustment factor (SBAF). The SBAF for each channel at different calibration sites is shown in Table III and Table IV.

TABLE III. SBAF LA3 AGAINST OL8 AND LA3 AT CALIBRATION SITE

Num	Sites	LA3-OL8			
		R	G	B	N
1	Libya 1	1,026	0,825	0,813	0,988
2	Mauritania 1	0,981	0,837	0,817	0,966
3	Mauritania 2	0,981	0,837	0,817	0,966
4	Libya 4	1,026	0,825	0,813	0,988
5	Algeria 3	1,026	0,825	0,813	0,988
6	La Crau	1,021	0,918	0,835	0,955

TABLE IV. SBAF LA3 AGAINST MS2 AND LA3 AT CALIBRATION SITE

Num	Sites	LA3-MS2			
		R	G	B	N
1	Libya 1	1,004	0,854	0,835	0,998
2	Mauritania 1	0,996	0,870	0,854	0,960
3	Mauritania 2	0,996	0,870	0,854	0,960
4	Libya 4	1,004	0,854	0,835	0,998
5	Algeria 3	1,004	0,854	0,835	0,998
6	La Crau	1,007	0,923	0,847	0,996

Based on Table III and Table IV, it can be seen that spectral differences between LA3 and OL8, as well as between LA3 and MS2, are not very significant, especially in the red and NIR channels. The SBAF results, which get close to 1 at a few calibration sites, make this clear. However, in the green channel (G), both LA3-OL8 and LA3-MS2 show lower SBAF values than the red and NIR channels. When considering each calibration site, La Crau has the best SBAF values in every channel compared to other sites.

After obtaining the SBAF values, the calibration coefficient calculation process is carried out. Calibration coefficients are derived using this procedure to convert the LA3 image's DN values into radiance. In this study, two types of calibration coefficients will be obtained: the calibration coefficient with OL8 as the reference and the calibration coefficient with MS2 as the reference. Fig. 4 and 5 display the calibration coefficients for each ROI at multiple calibration sites.

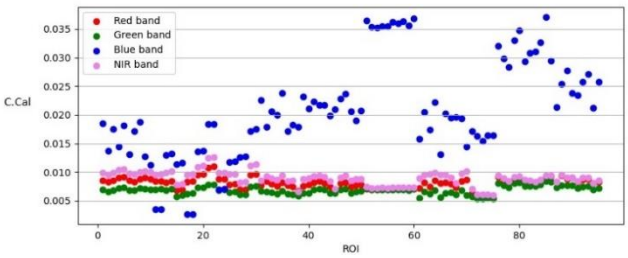


Fig. 4. Calibration coefficient of LA3 with reference OL8.

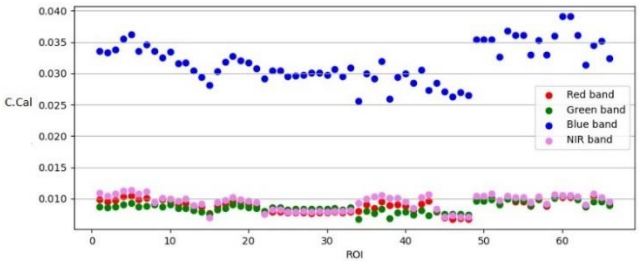


Fig. 5. Calibration coefficient of LA3 with reference MS2.

TABLE V. THE ROI NUMBERS FOR THE CALIBRATION SITES

Calibration of LA3 with Reference OL8		Calibration of LA3 with Reference MS2	
ROI	Calibration Site	ROI	Calibration Site
1-10	Libya 1	1-7	Libya 1
11-30	Algeria 5	8-17	Mauritania 1
31-40	Mauritania 1	18-22	Mauritania 2
41-50	Mauritania 2	23-33	Libya 4
51-60	Libya 4	34-43	Algeria 3
61-70	Algeria 3	44-48	La Crau
71-75	La Crau	49-61	Mauritania 1
76-85	Mauritania 1	62-66	Mauritania 2
86-95	Mauritania 2		

Fig. 4 shows the calibration coefficients of LA3 with OL8 as the reference, and Fig. 5 shows the calibration coefficients of LA3 with MS2 as the reference. Meanwhile, Table V shows the ROI numbers for the calibration sites used in Fig. 4 and 5. It is evident from both figures that the calibration coefficients vary across different calibration sites and different periods. If the averaging process is carried out based on on-site calibration and time, then a graph is obtained in Fig. 6 to 9.

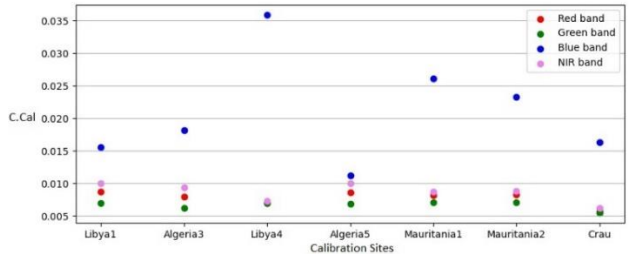


Fig. 6. Average LA3 calibration coefficient with OL8 reference at each calibration site.

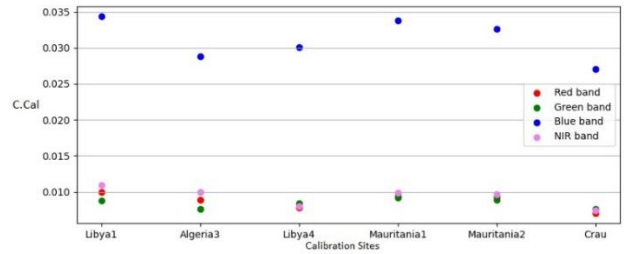


Fig. 7. Average LA3 calibration coefficient with MS2 reference at each calibration site.

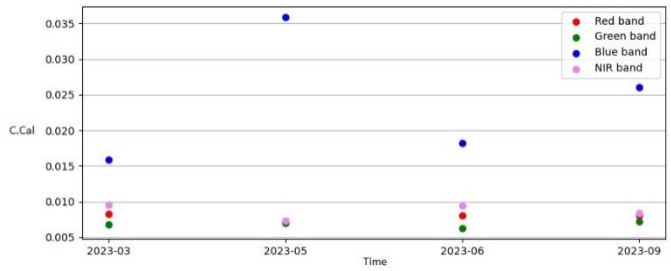


Fig. 8. Average LA3 calibration coefficient with OL8 reference based on time.

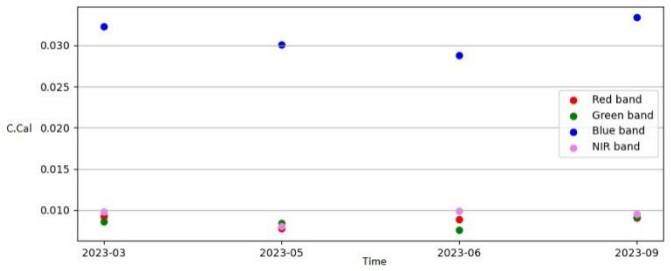


Fig. 9. Average LA3 calibration coefficient with MS2 reference based on time.

Fig. 6 shows the average calibration coefficient graph of LA3 with OL8 as the reference across different ROIs representing specific calibration sites. From the figure, it is evident that the red and NIR channels tend to remain stable in Libya 1, Algeria 3, Algeria 5, Mauritania 1, and Mauritania 2. The red channel ranges from 0.0080 to 0.0087, while the NIR channel ranges from 0.0087 to 0.0099. At Libya 4, the red and NIR channels are slightly lower than at other calibration sites in the Sahara Desert, but the blue channel is the highest compared to other sites. The calibration coefficients at this site are 0.007 for the red and green channels, 0.035 for the blue channel, and 0.0073 for the NIR channel. At La Crau, the red, green, and NIR channels are lower than at other calibration sites. Meanwhile, the calibration coefficient for the blue channel is lower than at other calibration sites, except for Libya 1. The calibration coefficients at this site are 0.0057 for the red channel, 0.0054 for the green channel, 0.016 for the blue channel, and 0.0062 for the NIR channel.

Fig. 7 shows the average calibration coefficient graph of LA3 with MS2 as the reference across different ROIs representing specific calibration sites. From the figure, it is evident that the red and NIR channels tend to remain stable in Libya 1, Algeria 3, Algeria 5, Mauritania 1, and Mauritania 2. The calibration coefficients range from 0.0088 to 0.0099 for the red channel and from 0.0091 to 0.011 for the NIR channel. The calibration coefficient for the green channel remains relatively stable across all calibration sites, ranging from 0.0076 to 0.0089. At Libya 4, the red and NIR channels are slightly lower than at other calibration sites in the Sahara Desert, with values of 0.0078 for the red channel and 0.0080 for the NIR channel. At La Crau, the red, green, blue, and NIR channels are lower than at other calibration sites. Meanwhile, the blue channel tends to be more stable than the calibration using OL8 as a reference, with values ranging from 0.026 to 0.034. This differs from the calibration using OL8 as a reference, where the blue channel is more fluctuating, with values ranging from 0.0080 to 0.035.

Fig. 8 shows the average calibration coefficient of LA3 with OL8 as the reference based on time. It can be observed that the red and green channel tends to remain stable in March, May, June, and September. In the NIR channel, there is a slight decrease in May and September compared to March and June. Meanwhile, the blue channel fluctuates, with values ranging from 0.015 to 0.035. Fig. 9 shows the average calibration coefficient of LA3 with OL8 as the reference over time. The figure indicates that the red and blue channels tend to be stable, with values ranging from 0.0078 to 0.0092 for the red channel and 0.028 to 0.032 for the blue channel. In March, there is a decrease in the calibration coefficient for the NIR channel, and in June, there is a decrease in the green channel.

C. Optimization Result

After obtaining the calibration coefficients, an optimization is carried out to derive a final calibration coefficient close to the reference value. The optimization is performed using a Kalman filter. Fig. 10 through 17 display the calibration coefficient optimization outcomes using the Kalman filter.

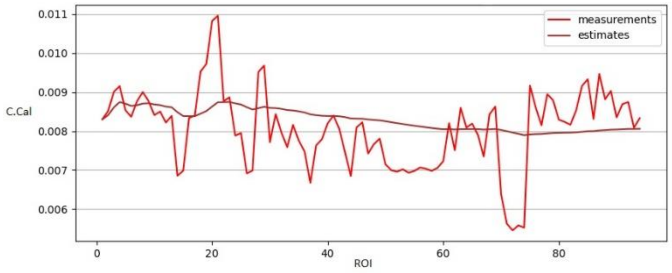


Fig. 10. Calibration coefficients optimization result of LA3 reference OL8 with one-dimensional Kalman filter on red channel.

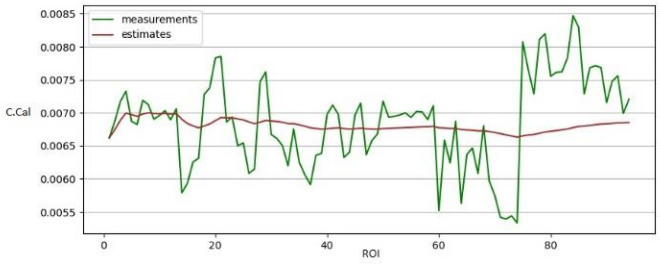


Fig. 11. Calibration coefficients optimization result of LA3 reference OL8 with one-dimensional Kalman filter on green channel.

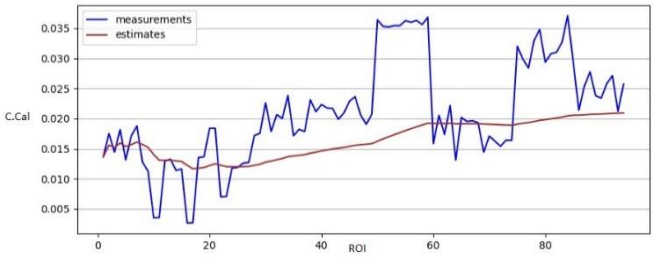


Fig. 12. Calibration coefficients optimization result of LA3 Reference OL8 with one-dimensional Kalman filter on blue channel.

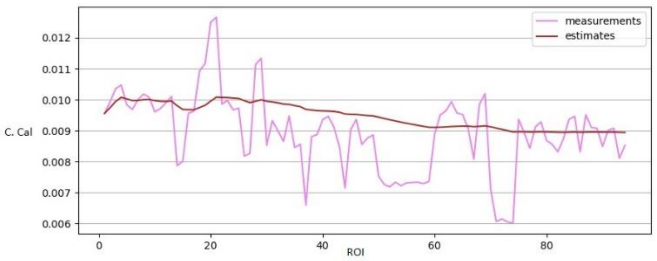


Fig. 13. Calibration coefficients optimization result of LA3 reference OL8 with one-dimensional Kalman filter on nir channel.

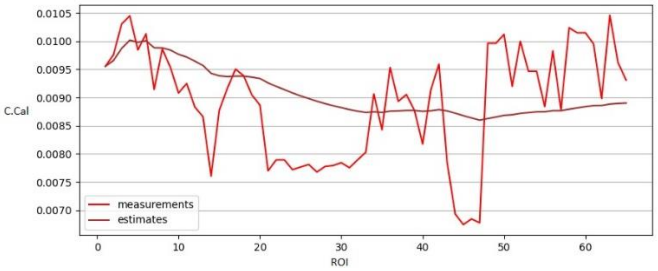


Fig. 14. Calibration coefficients optimization result of LA3 reference MS2 with one-dimensional Kalman filter on red channel.

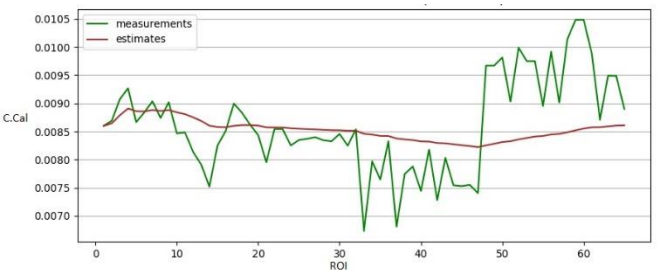


Fig. 15. Calibration coefficients optimization result of LA3 reference MS2 with one-dimensional Kalman filter on green channel.

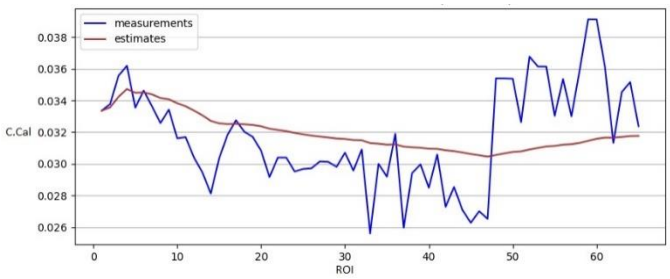


Fig. 16. Calibration coefficients optimization result of LA3 reference MS2 with one-dimensional Kalman filter on blue channel.

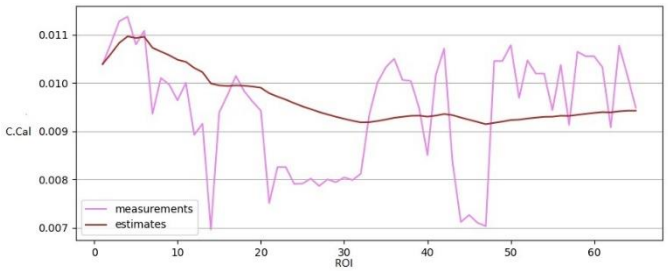


Fig. 17. Calibration coefficients optimization result of LA3 reference MS2 with one-dimensional Kalman filter on nir channel.

Fig. 10 to 13 show the comparison graphs of the LA3 calibration coefficients referenced to OL8 before and after optimization using the Kalman filter. It can be seen that these graphs experience a reduction in fluctuations, and the graphs resulting from the Kalman filter optimization tend to be more stable for each sample area. The final results of this optimization process are 0.0081 for the red channel, 0.0069 for the green channel, 0.02 for the blue channel, and 0.0089 for the NIR channel. Subsequently, these calibration coefficients will be cross-validated using MS2 imagery.

Meanwhile, Fig. 14 to 17 present the comparison graphs of the LA3 calibration coefficients referenced to MS2 before and after optimization using the Kalman filter. It can be observed that these graphs also show a reduction in fluctuations, and the graphs resulting from the Kalman filter optimization tend to be more stable for each sample area. The final results of this optimization process are 0.0089 for the red channel, 0.0086 for the green channel, 0.032 for the blue channel, and 0.0094 for the NIR channel. Subsequently, these calibration coefficients will be cross-validated using OL8 imagery.

D. Validation and Evaluation

The Digital Number (DN) can be converted into radiance using the calibration coefficients that were obtained during the calibration procedure. The first validation stage was performed using the same data and calibration sites as in the calibration process. However, the validation was carried out crosswise, i.e. validating LA3 radiance referenced to OL8 with MS2 radiance and validating LA3 radiance referenced to MS2 with OL8 radiance. This process was conducted on pre- and post-optimization data to assess the difference between LA3 radiance and the radiance of standardized satellites. The validation graphs of LA3 radiance with OL8 reference compared to MS2 radiance are shown in Fig. 18 to 21.

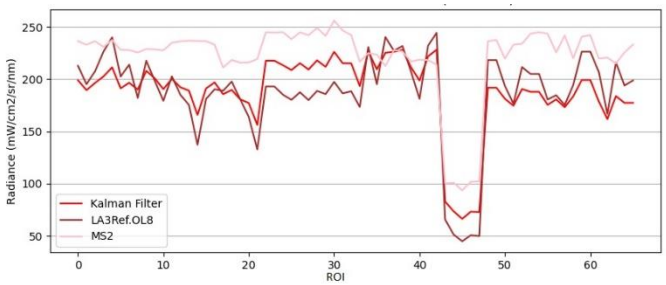


Fig. 18. Validation graph of LA3 radiance with OL8 reference before and after optimization on red channel compared to MS2 radiance.

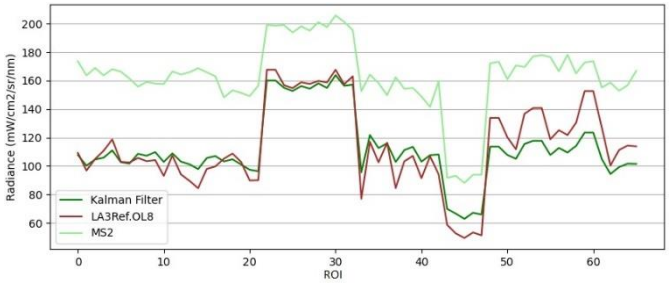


Fig. 19. Validation graph of LA3 radiance with OL8 reference before and after optimization on green channel compared to MS2 radiance.

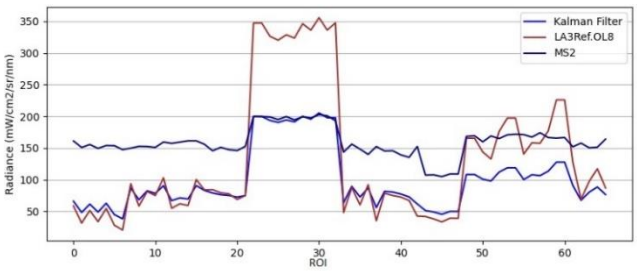


Fig. 20. Validation graph of LA3 radiance with OL8 reference before and after optimization on blue channel compared to MS2 radiance.

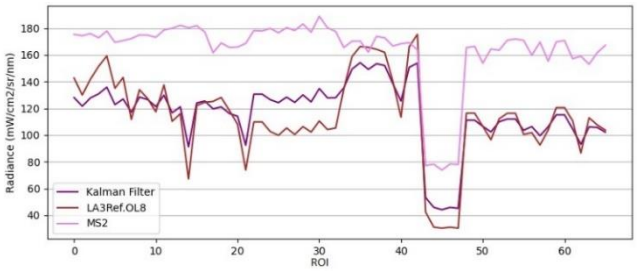


Fig. 21. Validation graph of LA3 radiance with OL8 reference before and after optimization on nir channel compared to MS2 radiance.

TABLE VI. PERCENTAGE DIFFERENCE IN LA3 RADIANCE WITH OL8 REFERENCE COMPARED TO MS2 RADIANCE

Sites	ROI	Percentage Difference in LA3 Radiance with OL8 Reference Compared to MS2 Radiance (%)							
		Before Optimization				After Optimization			
		R	G	B	N	R	G	B	N
Lib1	1-7	8.05	36.03	74.05	17.52	14.92	37.05	65.46	26.52
Mau1	8-17	20.44	39.72	50.58	34.20	17.19	35.25	49.52	32.29
Mau2	18-22	20.04	34.52	48.18	33.34	17.68	33.72	49.14	32.18
Lib4	23-33	23.34	18.99	70.31	41.54	12.11	20.88	0.72	28.71
Alg3	34-43	1.66	35.33	54.79	8,11	1.392	29.38	49.08	13.12
LCr	44-48	47.38	42.43	64.42	57,13	26.07	27.88	54.39	39.19
Mau12	49-61	14.60	23.07	3.58	34.01	21.08	33.68	33.73	34.44
Mau22	62-66	11.82	28.31	35.78	44.31	21.10	36.54	47.96	45.73

Fig. 18 to 21 show the validation graphs of LA3 with OL8 reference before and after optimization using the one-dimensional Kalman filter. MS2 radiance, the other reference camera, is used to compare radiance before and after optimization. The figures indicate that some lines get closer to the reference radiance line after optimization with the one-dimensional Kalman filter. However, some lines move further away after the optimization process. Based on Table VI, five sites in the red channel show a decrease in percentage difference, while three sites show an increase. In the green, blue, and NIR channels, four sites show a decrease in percentage difference, and four sites show an increase. Next, the radiance of LA3 with MS2 reference before and after optimization is compared to OL8 radiance. Fig. 22 to 25 are the validation graphs from this process.

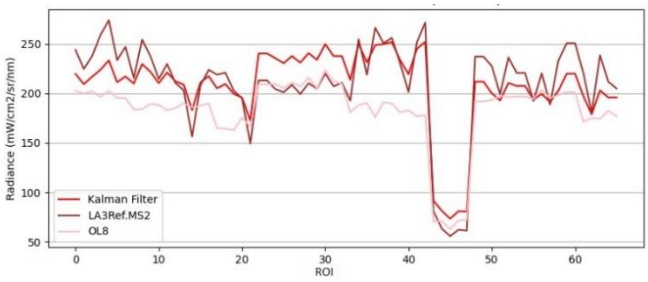


Fig. 22. Validation graph of LA3 radiance with MS2 reference before and after optimization on red channel compared to OL8 radiance.

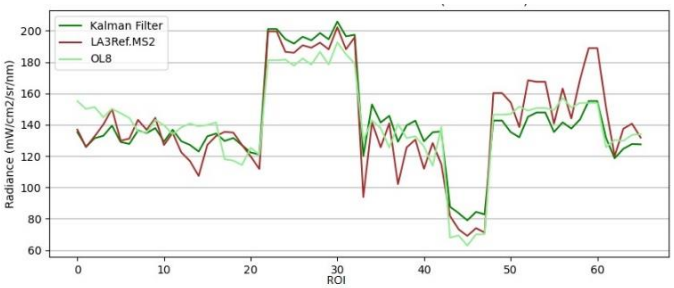


Fig. 23. Validation graph of LA3 radiance with MS2 reference before and after optimization on green channel compared to OL8 radiance.

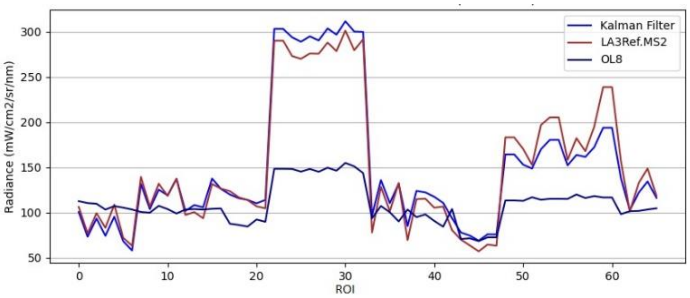


Fig. 24. Validation graph of LA3 radiance with MS2 reference before and after optimization on blue channel compared to OL8 radiance.

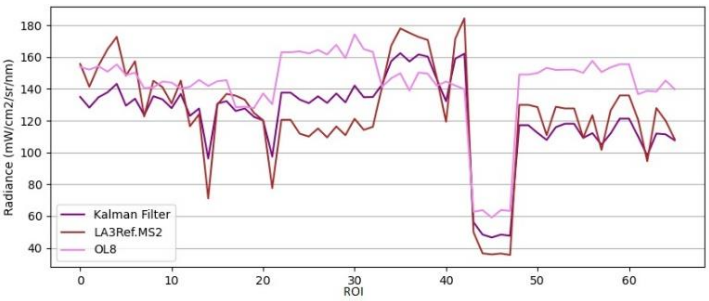


Fig. 25. Validation graph of LA3 radiance with MS2 reference before and after optimization on nir channel compared to OL8 radiance.

TABLE VII. PERCENTAGE DIFFERENCE IN LA3 RADIANCE WITH MS2 REFERENCE COMPARED TO OL8 RADIANCE

Sites	ROI	Percentage Difference in LA3 Radiance with MS2 Reference Compared to OL8 Radiance (%)							
		Before Optimization				After Optimization			
		R	G	B	N	R	G	B	N
Lib1	1-7	23.36	9.26	62.93	2.81	9.83	11.64	25.16	11.54
Mau 1	8-17	15.37	6.82	24.93	11.58	13.85	4.74	16.27	11.35
Mau 2	18-22	18.33	5.64	12.39	9.36	17.79	6.03	30.40	9.03
Lib4	23-33	1.08	5.70	128.42	29.94	12.85	8.36	101,90	17.77
Alg3	34-43	30.49	7.88	31.64	12.59	30.58	3.99	16.75	6.65
LCr	44-48	7.04	8.53	46.05	37.85	17.25	22.67	4.86	20.96
Mau 12	49-61	13.95	7.65	50.76	18.38	4.35	5.08	46.26	24.88
Mau 22	62-66	20.02	4.27	2.03	18.22	10.32	3.53	20.38	22.85

Fig. 22 to 25 show the validation graphs of LA3 with MS2 reference before and after optimization using the one-dimensional Kalman filter. The other reference camera, OL8, is used to compare the radiation levels before and after optimization. The figures indicate that some lines get closer to the reference radiance line after optimization with the one-dimensional Kalman filter. However, following the optimization procedure, certain lines get farther apart. Furthermore, the Libya-4 site exhibits a noteworthy percentage variation in the blue channel. Based on Table VII, in the red, green, and NIR channels, five sites show a decrease in percentage difference, while three sites show an increase in percentage difference. In the blue channel, six sites show a decrease in percentage difference, and two sites show an increase in percentage difference.

V. DISCUSSION

Based on the results, in the preprocessing stage of this study, digital numbers (DN) were derived for each image: LA3, OL8, and MS2. Subsequently, SBAF calculations were performed. The results show that the SBAF values in the RGBN channels are close to 1 (Tables III and IV). This indicates that the spectral response of LA3 is similar to that of the reference cameras (OL8 and MS2). Once the SBAF values were obtained, calibration coefficients could be calculated using DN, SBAF, reflectance, and AOD. From this process, calibration coefficients are shown in Fig. 4 and 5. The graph indicates that the blue channel's calibration coefficient is higher than the other channels. This is because the digital number for the blue channel is lower than the other channels, requiring a high calibration coefficient to reach the reference radiance value.

The calibration coefficients obtained during the calculation process showed fluctuations across each ROI, whereas, in theory, they should have similar values at the calibration site. Therefore, this study used a Kalman Filter to reduce noise in the calibration coefficients. Furthermore, the Kalman Filter was employed to optimize the calibration coefficients, resulting in a single calibration coefficient value. These values are used to convert DN to TOA radiance.

Based on the results, the optimization process using the Kalman Filter reduce noise in the calibration coefficient graphs (Fig. 10 until Fig. 17). This can be seen from the calibration coefficient graphs, which initially showed fluctuations but became more stable for each ROI. The calibration coefficients for LA3 with OL8 and MS2 reference were obtained from the optimization. The calibration coefficient for LA3 with OL8 reference is 0.0081 for the red channel, 0.0069 for the green channel, 0.02 for the blue channel, and 0.0089 for the NIR channel. The calibration coefficient of LA3 with MS2 reference are 0.0089 for the red channel, 0.0086 for the green channel, 0.032 for the blue channel, and 0.0094 for the NIR channel. Next, these calibration coefficients were validated using another reference camera as the validator. The calibration coefficient for LA3 with OL8 reference was validated with MS2, while the calibration coefficient for LA3 with MS2 reference was validated with OL8.

The cross-validation results show that the LA3 calibration process with MS2 as the reference results in radiance values closer to the reference radiance than OL8 as the reference. This implies that using MS2 as the reference camera is more sensible for the LA3 cross-calibration procedure. This calibration technique demonstrates that, for most images, the red, green, blue, and NIR channels can be effectively optimized with a one-dimensional Kalman filter. However, further development is needed to achieve better results. Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF), or other methods can be considered for further research so that the radiance produced from the cross-calibration process more closely matches the reference radiance values.

VI. CONCLUSION

The multi-site cross-calibration method with one-dimensional Kalman filter optimization can be implemented on the LA3 camera. It is possible to optimize calibration coefficients using the one-dimensional Kalman Filter, as can be inferred from the discussion in the preceding section. This is evident from the reduction in the percentage difference between LA3 radiance and reference radiance after the optimization process. The results also show that LA3 calibration process with MS2 as the reference results in radiance values closer to the reference radiance than OL8 as the reference. This implies that using MS2 as the reference camera is more sensible for the LA3 cross-calibration procedure. Basically, this method can be applied on LA3. However, further development is needed to achieve better results so that the radiance produced from the cross-calibration process more closely matches the reference radiance values.

ACKNOWLEDGMENT

The author would like to thank Mr. Mohammad Mukhayadi and Mr. Patria Rachman Hakim as group leader and leader satellite operation missions of Satellite Technology Research Center BRIN, for their support and assistance so that this works can be well completed.

REFERENCES

[1] P. R. Hakim, A. H. Syafrudin, S. Salaswati, S. Utama, and W. Hasbi, "Development of Systematic Image Preprocessing of LAPAN-A3/IPB Multispectral Images," vol. 7, no. 10, pp. 9–18, 2019, [Online]. Available: http://arxiv.org/abs/1901.09189.

[2] S. Bolívar-Santamaría and B. Reu, "Detection and characterization of agroforestry systems in the Colombian Andes using sentinel-2 imagery," *Agrofor. Syst.*, vol. 95, no. 3, pp. 499–514, 2021, doi: 10.1007/s10457-021-00597-8.

[3] M. T. Parveen and R. A. Ilahi, "Assessment of land-use change and its impact on the environment using GIS techniques: a case of Kolkata Municipal Corporation, West Bengal, India," *GeoJournal*, vol. 87, no. s4, pp. 551–566, 2022, doi: 10.1007/s10708-022-10581-z.

[4] H. Fadaei, "Advanced land observing satellite data to identify ground vegetation in a juniper forest, northeast Iran," *J. For. Res.*, vol. 31, no. 2, pp. 531–539, 2020, doi: 10.1007/s11676-018-0812-5.

[5] A. Sharifi, S. Felegari, and A. Tariq, "Mangrove forests mapping using Sentinel-1 and Sentinel-2 satellite images," *Arab. J. Geosci.*, vol. 15, no. 20, 2022, doi: 10.1007/s12517-022-10867-z.

[6] C. C. Liu, Y. H. Chen, M. H. M. Wu, C. Wei, and M. H. Ko, "Assessment of forest restoration with multitemporal remote sensing imagery," *Sci. Rep.*, vol. 9, no. 1, pp. 1–19, 2019, doi: 10.1038/s41598-019-43544-5.

[7] M. Tamang, S. Nandy, R. Srinet, A. K. Das, and H. Padalia, "Bamboo Mapping Using Earth Observation Data: A Systematic Review," *J. Indian Soc. Remote Sens.*, vol. 50, no. 11, pp. 2055–2072, 2022, doi: 10.1007/s12524-022-01600-0.

[8] A. D. Prasad, P. Ganasala, R. Hernández-Guzmán, and F. Fathian, "Remote sensing satellite data and spectral indices: an initial evaluation for the sustainable development of an urban area," *Sustain. Water Resour. Manag.*, vol. 8, no. 1, pp. 1–16, 2022, doi: 10.1007/s40899-022-00607-2.

[9] A. K. Wijayanto, S. M. Yusuf, and W. A. Pambudi, "The Characteristic of spectral reflectance of LAPAN-IPB (LAPAN-A3) Satellite and Landsat 8 over agricultural area in Probolinggo, East Java," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 284, no. 1, 2019, doi: 10.1088/1755-1315/284/1/012004.

[10] L. F. Amalo, I. A. Nur, and N. R. Rochimawati, "Drought monitoring using LISAT (LAPAN-IPB Satellite) and Landsat 8 Satellite Imagery in Pakisjaya District, West Java," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 284, no. 1, 2019, doi: 10.1088/1755-1315/284/1/012008.

[11] P. A. Permatasari, S. Muslimah, and B. A. Utomo, "Comparison of LISAT and Landsat imagery for estimating chlorophyll-a (case study: Jatiluhur Reservoir)," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 284, no. 1, 2019, doi: 10.1088/1755-1315/284/1/012041.

[12] M. A. Raimadoya, B. H. Trisasongko, and A. Zain, "Analisis Misi Dan Rancangan Lapan-Ipb Satellite (Lisat) Untuk Pemantauan Kemandirian Pangan," *J. Ilmu Pertan. Indones.*, vol. 16, no. 3, pp. 173–178, 2011.

[13] S. P. Nugroho, L. D. W. Handayani, R. Meidiza, and G. Munggaran, "Landuse change analysis for hydrology response and planning management of Cibeet Sub-Watershed, West Java, Indonesia," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 284, no. 1, 2019, doi: 10.1088/1755-1315/284/1/012002.

[14] Q. Hao *et al.*, "The Development of Snapshot Multispectral Imaging Technology Based on Artificial Compound Eyes," *Electron.*, vol. 12, no. 4, 2023, doi: 10.3390/electronics12040812.

[15] Q. Song *et al.*, "Time Series Analysis-Based Long-Term Onboard Radiometric Calibration Coefficient Correction and Validation for the HY-1C Satellite Calibration Spectrometer," *Remote Sens.*, vol. 14, no. 19, 2022, doi: 10.3390/rs14194811.

[16] K. Arai, "Comparison Among Cross, Onboard and Vicarious Calibrations for Terra/ASTER/VNIR," *Int. J. Adv. Res. Artif. Intell.*, vol. 2, no. 10, pp. 14–18, 2013, doi: 10.14569/ijarai.2013.021003.

[17] S. Zhu *et al.*, "In-Flight Relative Radiometric Calibration of a Wide Field of View Directional Polarimetric Camera Based on the Rayleigh Scattering over Ocean," *Remote Sens.*, vol. 14, no. 5, pp. 1–19, 2022, doi: 10.3390/rs14051211.

[18] S. Zhu *et al.*, "Evaluation of radiometric performance of MODIS visible bands using the Rayleigh scattering method," *J. Appl. Remote Sens.*, vol. 13, no. 01, p. 1, 2019, doi: 10.1117/1.jrs.13.018503.

[19] X. Chen *et al.*, "In-flight calibration of GF-1/WFV visible channels using Rayleigh scattering," *Remote Sens.*, vol. 9, no. 6, 2017, doi: 10.3390/rs9060513.

[20] K. Tan *et al.*, "Vicarious Calibration for the AHSI Instrument of Gaofen-5 with Reference to the CRCS Dunhuang Test Site," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3409–3419, 2021, doi: 10.1109/TGRS.2020.3014656.

[21] B. Alhammoud *et al.*, "Sentinel-2 Level-1 Radiometry Assessment Using Vicarious Methods from DIMITRI Toolbox and Field Measurements from RadCalNet Database," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 9, pp. 3470–3479, 2019, doi: 10.1109/JSTARS.2019.2936940.

[22] K. N. Babu, N. Kumawat, and M. R. Pandya, "Radiometric Calibration of AVIRIS-NG sensor using Indian desert sites," *Adv. Sp. Res.*, no. xxxx, 2023, doi: 10.1016/j.asr.2023.01.022.

[23] H. Tang, J. Xie, W. Chen, H. Zhang, and H. Wang, "Absolute Radiometric Calibration of ZY3-02 Satellite Multispectral Imager Based on Irradiance-Based Method," 2023.

[24] J. Lu, T. He, S. Liang, and Y. Zhang, "An Automatic Radiometric Cross-Calibration Method for Wide-Angle Medium-Resolution Multispectral Satellite Sensor Using Landsat Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2022, doi: 10.1109/TGRS.2021.3067672.

[25] C. Gao *et al.*, "Radiometric Cross-Calibration of GF-4 / VNIR Sensor," vol. 13, pp. 2337–2350, 2020.

[26] H. Gao, "Cross-Calibration of GF-1 PMS Sensor With," vol. 54, no. 8, pp. 4847–4854, 2016.

[27] J. Han, Z. Tao, Y. Xie, Q. Liu, and Y. Huang, "Radiometric Cross-Calibration of GF-4/PMS Based on Radiometric Block Adjustment," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 4522–4534, 2021, doi: 10.1109/TGRS.2020.3009740.

[28] H. Mizuochi, S. Tsuchida, K. Obata, H. Yamamoto, and S. Yamamoto, "Combination of cross-and inter-band radiometric calibrations for a hyperspectral sensor using model-based spectral band adjustment," *Remote Sens.*, vol. 12, no. 12, 2020, doi: 10.3390/rs12122011.

[29] K. Arai, "Vicarious Calibration Based Cross Calibration of Solar Reflective Channels of Radiometers Onboard Remote Sensing Satellite and Evaluation of Cross Calibration Accuracy through Band-to-Band Data Comparisons," *Int. J. Adv. Comput. Sci. Appl.*, vol. 4, no. 3, pp. 7–14, 2013, doi: 10.14569/ijacsa.2013.040312.

[30] A. H. Syafrudin, S. Salaswati, and W. Hasbi, "Pre-Flight Radiometric Model of Linear Imager on LAPAN-IPB Satellite," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 149, no. 1, 2018, doi: 10.1088/1755-1315/149/1/012068.

[31] K. Arai *et al.*, "Method for uncertainty evaluation of vicarious calibration of spaceborne visible to near infrared radiometers," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 1, pp. 387–393, 2019, doi: 10.14569/IJACSA.2019.0100151.

[32] S. Salaswati, P. R. Hakim, A. H. Syafrudin, R. Hartono, and S. Utama, "Vicarious Radiometric Calibration of Lapan-A3 / IPB Satellite Multispectral Imager in Jaddih Hill Madura," *J. Aerosp. Technol.*, pp. 31–42, 2020.

[33] S. Salaswati, P. R. Hakim, A. H. Syafrudin, R. Hartono, and S. Utama, "Analysis of the Atmospheric Effects on the Vicarious Calibration of LAPAN-A3/IPB Satellite Multispectral Camera in Jaddih Hill Madura," *J. Aerosp. Technol.*, vol. 20, no. 1, pp. 55–67, 2022.

[34] C. Bacour, X. Briottet, F. M. Bréon, F. Viallefont-Robinet, and M. Bouvet, "Revisiting Pseudo Invariant Calibration Sites (PICS) over sand deserts for vicarious calibration of optical imagers at 20 km and 100 km scales," *Remote Sens.*, vol. 11, no. 10, pp. 1–28, 2019, doi: 10.3390/rs11101166.

[35] CEOS, "Radiometric Calibration Network Portal." https://www.radcalnet.org/#!/.

[36] C. Revel *et al.*, "Sentinel-2A and 2B absolute calibration monitoring," *Eur. J. Remote Sens.*, vol. 52, no. 1, pp. 122–137, 2019, doi: 10.1080/22797254.2018.1562311.

[37] M. A. Ridwan *et al.*, "Applications of landsat-8 data: A Survey," *Int. J. Eng. Technol.*, vol. 7, no. 4, pp. 436–441, 2018, doi: 10.14419/ijet.v7i4.35.22858.USGS, "What are the best Landsat spectral bands for use in my research?" https://www.usgs.gov/faqs/what-are-best-landsat-spectral-bands-use-my-research (accessed Jun. 10, 2023).

[38] E. S. Agency, *Sentinel-2 User Handbook*. ESA Standard Document, 2015.

[39] L. Liu, T. Shi, H. Gao, X. Zhang, Q. Han, and X. Hu, "Long-term cross calibration of HJ-1A CCD1 and Terra MODIS reflective solar bands," *Sci. Rep.*, vol. 11, no. 1, pp. 1–14, 2021, doi: 10.1038/s41598-021-86619-y.

Combining BERT and CNN for Sentiment Analysis A Case Study on COVID-19

Gunjan Kumar[1], Renuka Agrawal[2], Kanhaiya Sharma[3], Pravin Ramesh Gundalwar[4], Aqsa kazi[5], Pratyush Agrawal[6], Manjusha Tomar[7] , Shailaja Salagrama[8]

Department of Computer Science & Engineering, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India[1, 2, 3, 5]

Department of Information Technology, Anurag University, Hyderabad India[4]

Department of Artificial Intelligence and Machine Learning, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India[6]

Basic Engineering Department, Indira College of Engineering and Management Pune, India[7]

Computer Information System, University of the Cumberland's, Williamsburg, Kentucky, USA[8]

***Abstract*—This research focuses on sentiment analysis to understand public opinion on various topics, with an emphasis on COVID-19 discussions on Twitter. By utilizing state-of-the-art Machine Learning (ML) and Natural Language Processing (NLP) techniques, the study analyzes sentiment data to provide valuable insights. The process begins with data preparation, involving text cleaning and length filtering to optimize the dataset for analysis. Two models are employed: a Bidirectional Encoder Representations from Transformers (BERT)-based Deep Learning (DL) model and a Convolutional Neural Network (CNN). The BERT model leverages transfer learning, demonstrating strong performance in sentiment classification, while the CNN model excels at extracting contextual features from the input text. To further enhance accuracy, an ensemble model integrates predictions from both approaches. The study emphasizes the ensemble technique's value for more precise sentiment analysis. Evaluation metrics, including accuracy, classification reports, and confusion matrices, validate the effectiveness of the proposed models and the ensemble approach. This research contributes to the growing field of social media sentiment analysis, particularly during global health crises like COVID-19, and underscores its potential to aid informed decision-making based on public sentiment.**

***Keywords*—Sentiment analysis; COVID-19; BERT; CNN; ensemble model; NLP; transfer learning**

I. Introduction

The global impact of the COVID-19 epidemic has shown a surge in communication and information sharing on social media platforms [1]. Notably, Twitter sentiment has emerged as a vital channel for individuals to express their ideas, concerns, and opinions regarding the pandemic. Understanding the sentiment conveyed in these tweets is crucial, as it provides valuable insights into the public's emotional responses [2, 3], concerns, and expectations related to the ongoing health crisis. This study embarks on a sentiment analysis journey, aiming to decipher the emotion and tone of tweets about COVID-19. Sentiment analysis has become indispensable for grasping how society perceives and responds to emerging events, such as the COVID-19 epidemic [4]. Analyzing tweet sentiments enables the identification of patterns in public opinion, tracking changes in public mood over time, and evaluating public satisfaction with governmental responses. In the era of rapid digital transformation, the global influence of social language cannot be overlooked. Platforms such as social media, Blogs, Online forums, and product review portals have become modern channels for expressing opinions, collectively forming a vast sea of User Generated Content that significantly impacts real-world businesses [5]. Analyzing public sentiment in these social media data entails Opinion Mining, deciphering the emotional intent behind comments or tweets – positive, negative, or neutral. This process, guided by patterns identified through text mining, enables the prediction of subsequent textual threads. Sentiment analysis also plays a crucial role in shaping a brand's presence on social media by providing inferences of an in-depth understanding of the target audience's perspective, Evaluation of competitors' emotional marketing, campaigns, responses, and identification of industry trends for strategic brand positioning and marketing [6].

Numerous studies have explored ML and DL models for sentiment analysis, particularly in the context of COVID-19-related tweets. The research in [7, 8] underscores the necessity and significance of sentiment analysis across various domains. Naïve Bayes classifier achieved an accuracy of 79% on a dataset comprising 11,974 COVID-19 tweets [9]. A comparative analysis of various ML models demonstrated their efficacy in COVID19 Tweets [10]. BERT outperformed traditional models such as Logistic Regression and Support Vector Machines (SVM), attaining an accuracy of 89% in the analysis of Indian tweets; however, the study was limited by its geographic focus and a lack of comprehensive dataset evaluation [11]. Furthermore, BERT exhibited high accuracy on Twitter datasets from Amazon and Hachette, though it did not thoroughly address biases inherent in social media data [12]. An RNN model surpassed SVM in global sentiment analysis, yet failed to consider data transparency and ethical implications [13]. In another study, fuzzy logic combined with deep learning yielded an accuracy of 81%, but lacked generalized recommendations [14]. Additionally, the Extra Trees Classifier demonstrated superior accuracy relative to other ML models [15]. Comparative studies in [16] and [17] emphasized ensemble techniques while acknowledging

challenges in generalizing results to non-English datasets of COVID-19 tweets.

The research gap highlights the need for hybrid models that integrate ML and DL approaches, aiming to enhance accuracy and address biases in sentiment analysis across diverse datasets.

This study utilizes advanced ML and NLP techniques for this purpose, and to preprocess the Twitter dataset effectively, by eliminating extraneous data, links, and special characters, excluding tweets with insufficient text content to generate a focused and insightful dataset. Subsequently, we explore two distinct modeling strategies. The first leverages BERT, a DL model recognized for its exceptional ability to capture context and semantics due to pre-training on extensive text data. The second method involves a CNN, a prominent ML model, to extract local text characteristics. These models are analyzed based on accuracy, utilizing various categorization criteria to determine their effectiveness. Additionally, we employ an ensemble strategy that merges predictions from both models to enhance overall accuracy. Ensemble modeling leverages the strengths of each model, which is known to strengthen outcomes while mitigating individual model drawbacks. In the ongoing COVID-19 outbreak, this study aims to illuminate public sentiment on social media. Decision-makers, public health experts, and politicians can utilize these insights to stay informed about evolving public opinions and concerns related to the pandemic. The objectives of this study are, evaluate the effectiveness of BERT and CNN models in classifying COVID-19 Twitter sentiment using accuracy, classification reports, and confusion matrices, and enhance sentiment analysis accuracy through an ensemble technique combining BERT and CNN predictions, providing more reliable results.

The rest of the manuscript is organized. Section II details the ML models used for analysis. Section III describes the proposed work, dataset, and methodology. Section IV presents the results, and Section V concludes with future directions.

II. Sentiment Analysis

This study employs BERT, CNN, and ensemble models to conduct sentiment analysis on COVID-19-related tweets. By leveraging BERT's contextual understanding in combination with the feature extraction strengths of CNN, the ensemble approach enhances both accuracy and robustness. The insights generated aim to deepen understanding of public sentiment during the pandemic, offering valuable information for decision-makers.

A. Bidirectional Encoder Representations from Transformers

Natural language processing has seen a revolutionary development, primarily due to the advancements made possible by deep learning and advanced neural network architectures. Among these developments, the pre-trained transformer-based model BERT has been crucial in changing the field of NLP applications [18]. In contrast to conventional unidirectional models, BERT's bidirectional approach considers both words that come before and after to grasp the context of a sentence. This fundamental change allows BERT to gain a deep comprehension of context, which helps it become skilled at managing linguistic subtleties. With massive corpora of pre-trained text, BERT is proficient in understanding syntactic structures and semantic links, producing rich contextual representations that provide a solid basis for a range of NLP applications. BERT is pre-trained on large text data using two tasks: masked Language Modeling and next-sentence prediction [19]. One noteworthy application of BERT is sentiment analysis, demonstrating its ability to identify the dominant sentiment or mood in each text [20, 21]. Sentiment analysis models with BERT can refine accuracy and granularity by utilizing rich contextual embedding's to decode minor variations in sentiment within the textual content, and architecture for tweets input and output is depicted in Fig. 1.

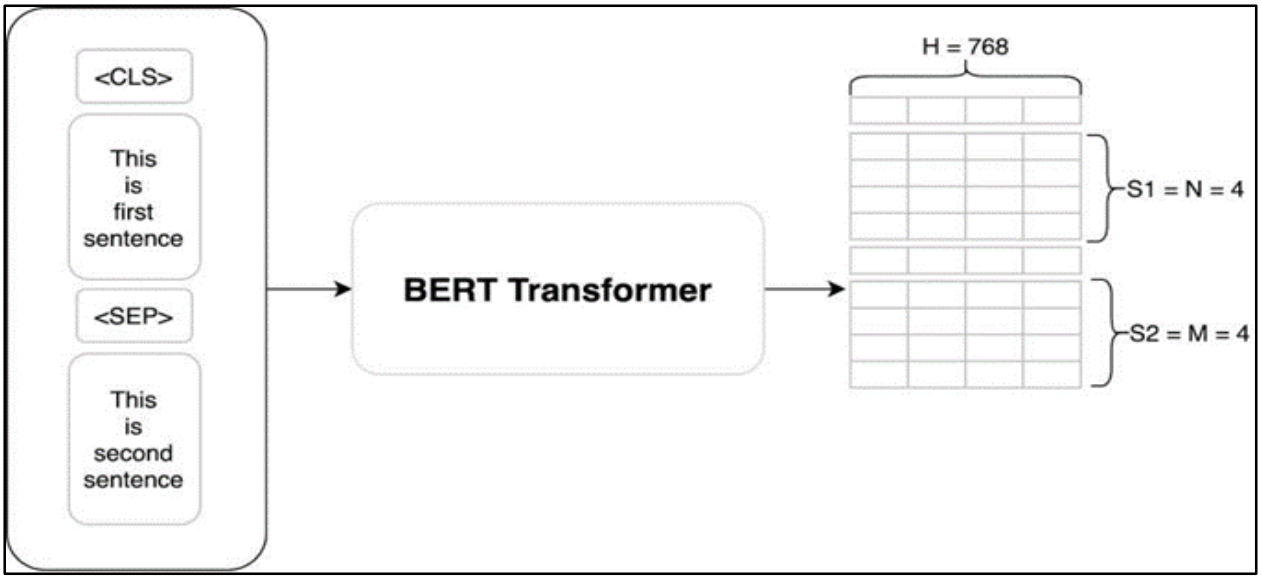


Fig. 1. BERT architecture for tweets input and output.

B. Convolutional Neural Network

The CNN-based COVID-19 sentiment analysis model architecture includes layers for token embedding, convolution, max-pooling, and fully connected neural networks [22]. Convolutional layers extract local text features, and max-pooling determines standout features, enabling CNN to effectively comprehend complex sentiment within tweets. CNNs excel at automatically learning and extracting features from structured data. It consists of four layers: the convolutional layer, pooling layer, fully connected layer, and output layer, as shown in Fig. 2. The convolutional layer includes convolutions where filters are applied to extract useful information and post-convolution, a Rectified Linear Unit (ReLU) function is used to introduce non-linearity. Pooling layers reduce the spatial dimensions while retaining significant information. Next, the information from previous layers is flattened into a one-dimensional vector in fully connected layers. This is further processed in dense layers, which combine the extracted features to make the final predictions. The final layer in CNN is the output layer, which typically uses a softmax function (for classification tasks) or another activation function depending on the needed task.

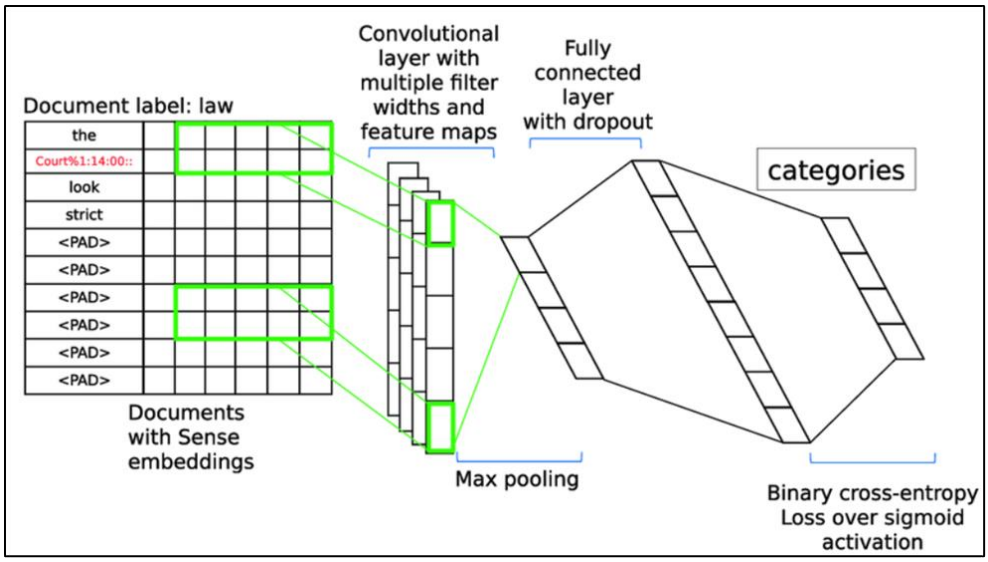


Fig. 2. CNN architecture [23].

C. Ensemble Model

Combining multiple models' advantages, ensemble models considerably improve prediction accuracy, resilience, and generalization. They are essential to many high-achieving machine learning systems, especially in challenging and vital applications. To improve COVID-19 sentiment analysis, the ensemble model shown in Fig. 3 combines predictions from the BERT model with the CNN. Through the process of average outcomes, the ensemble makes the most of the unique qualities of each model, so increasing the overall efficacy and reliability of sentiment classification. By removing individual model limitations, this ensemble technique fully understands public sentiment [24].

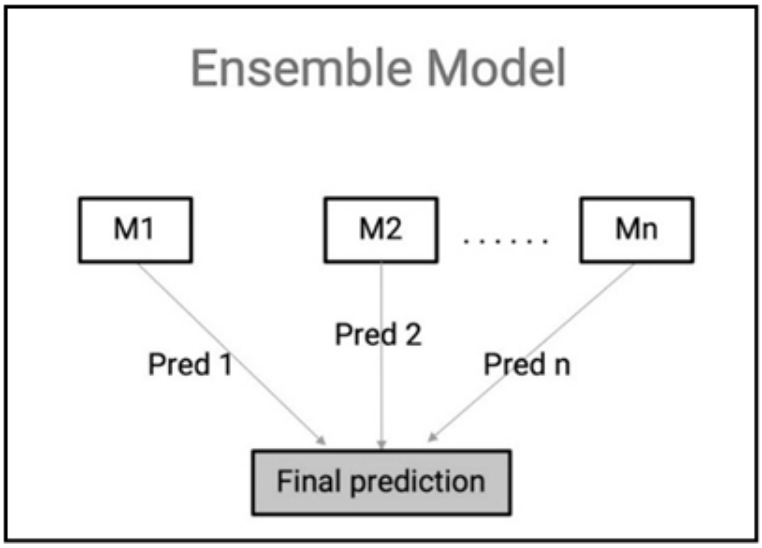


Fig. 3. Ensemble model.

III. METHODOLOGY

The suggested methodology outlines the steps involved in applying sentiment analysis to a corpus of tweets about COVID-19 and categorizing them as "Negative," "Neutral," or "Positive." It includes a series of crucial procedures intended to suitably set up, model, and assess the sentiment analysis. The steps include data preprocessing, text vectorization, machine learning, and deep learning model-based prediction. Furthermore, an ensemble model is constructed, which improves sentiment classification accuracy by combining the benefits of two different and efficient methods. The approach offers insights into public opinion on social media regarding the COVID-19 epidemic by showcasing the application of well-liked models and techniques for sentiment analysis on textual data for IMDB review dataset of 25000 samples [25]. This systematic process guarantees a thorough strategy for insightful analysis and outcomes.

A. Library Imports

The code begins by importing the necessary Python libraries for DL like Tensor Flow and Keras, ML libraries like Scikit-learn, and for data manipulation NumPy, and Pandas, for NLP NLTK, and for data visualization Matplotlib and Seaborn have been used. Furthermore, libraries such as Transformers for BERT- based models and Emoji for handling emojis are imported and installed.

B. Data Loading and Exploration

In this study, COVID-19 dataset consists of tweets collected over a 15-day period from March 16 to March 31, 2020. Twitter samples were used to analyze the trend and sentiment of COVID-related topics. Mounting Google Drive to access the dataset is how the code starts. Pandas Data Frames are loaded with two CSV files. df, which contains training data, and df_test, which contains testing data. To perform exploratory data analysis and display tweet counts by date and location, data pretreatment entails deleting duplicates, changing the 'TweetAt' field to date format, and more, as shown in Fig. 4 which shows that the tweet count by Date for half of a specific Month, and number of tweets by people across the Globe are considered for a complete month for a realistic analysis of their sentiments during Pandemic. Fig. 5 shows tweets across the Globe by different citizens; 15 states across the USA and Europe have been considered. Fig. 6 and Fig. 7 show a graphical analysis of COVID-19 tweets based on the length of tweets. Fig. 6 analyses of tweets with less than 10 words for the training dataset, whereas Fig. 7, represents a graphical representation of tweets with a length of less than 10 words with test data.

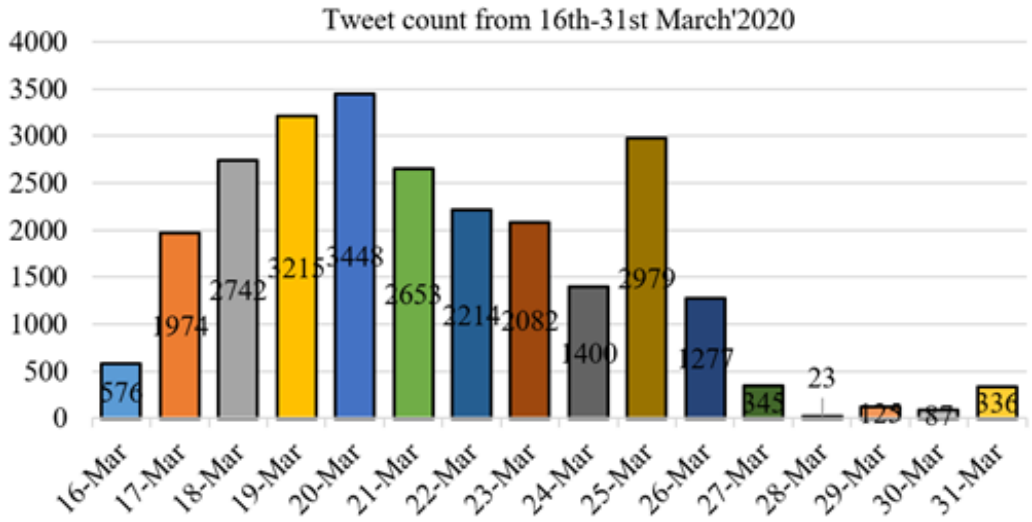


Fig. 4. Tweet count by date.

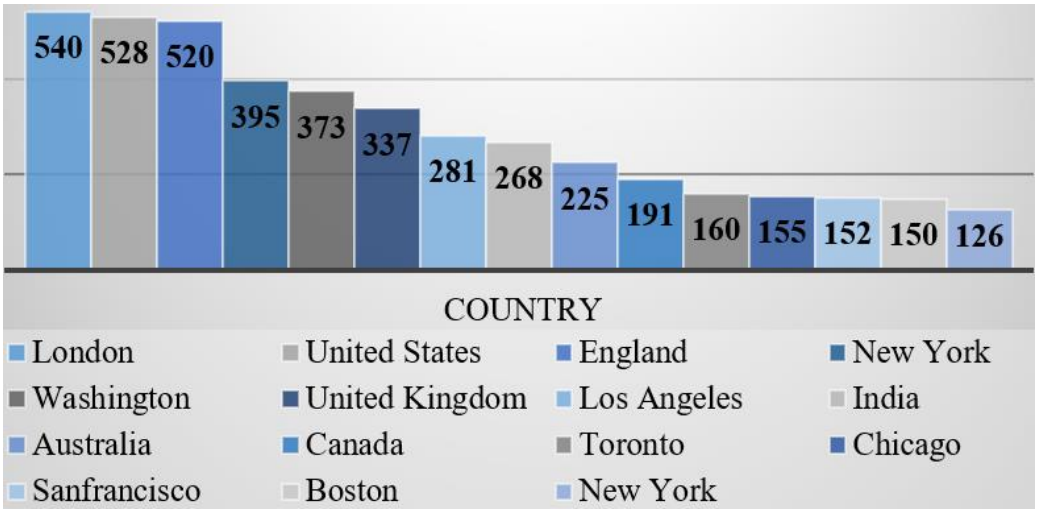


Fig. 5. Tweet count by country.

C. Data Preparation

In this study, Data Preparation is done using four stages: Text Preparation, Text Length analysis, Data balancing, and Text Vectorization. In the initial stage of Text preparation, Custom functions are employed for cleaning and preparing text data. These functions eliminate emojis, punctuation, links, mentions, and special characters, as well as handle hashtags. The pre-processed text is then saved in the Data Frames as new columns. In Text Length analysis the code computes the length in number of words of each pre-processed tweet and visualizes the tweet length distribution. The resulting histograms are illustrated in Fig. 6 and Fig. 7. Neural networks and machine learning rely heavily on data, but the quality of that data determines how well a model performs. Data science and data mining initiatives frequently deal with imbalanced datasets, which are those in which one class called a minority class has a comparatively smaller number of instances than the other classes. In preparing the training dataset, to address the class imbalance in sentiment labels using the RandomOverSampler from the imbalanced-learn library, resulting in a total of 15,148 samples. In the final stage of text preparation, authors vectorize the data using Count Vectorizer and Tfidf Transformer, converting textual information into numerical features suitable for training machine learning models. The Count Vectorizer transforms text data into a count matrix, which is then normalized into a TF or TF-IDF representation. Here, TF refers to term frequency, while TF-IDF combines term frequency with inverse document frequency. This widely used term weighting scheme is effective in information retrieval and document classification. For the proposed approaches, 3,787 data samples were utilized to evaluate model performance. The cleaned text is visualized in Fig. 8.

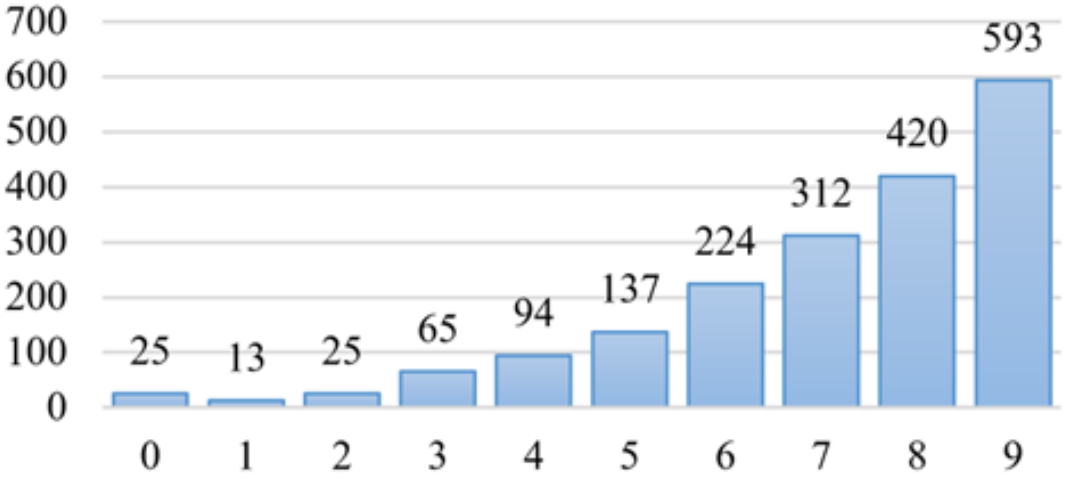


Fig. 6. The training dataset for COVID-19 tweets with a length of less than 10 words.

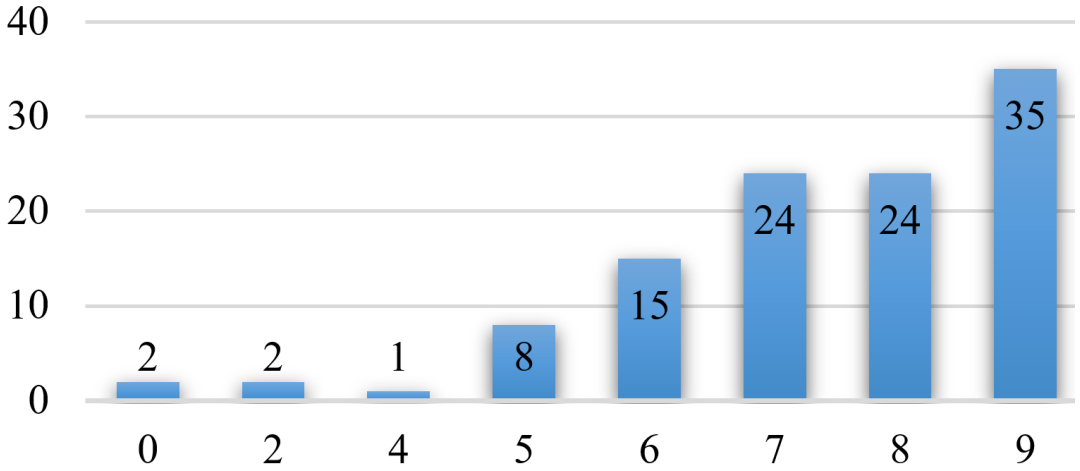


Fig. 7. Testing dataset for COVID-19 Tweets with length less than 10 words.

	OriginalTweet	Sentiment	text_clean	text_len	token_lens
3505	Stop misusing ur privilege amp grow up Some1 c...	Positive	stop misusing ur privilege amp grow up some1 c...	57	73
1789	For those that are cashlong, patient,calm&...	Extremely Positive	for those that are cashlong patientcalmamphave...	44	71
855	Lidl is total chaos, queues as long as the ais...	Extremely Negative	lidl is total chaos queues as long as the aisl...	62	70

Fig. 8. Text vectorization of the tweets.

D. Model Preparation

The algorithm optimizes a sentiment analysis pre-trained BERT model using the Hugging Face Transformers library. Text data is tokenized and transformed into BERT model input features. The model is trained on oversampled data and tested on new data, with a confusion matrix and classification metrics computed. Using a CNN model for sentiment analysis is defined and trained, incorporating an embedding layer, convolutional layers, and fully linked layers. The CNN model is tested based on the test data, and its accuracy is estimated.

Predictions from the BERT and CNN models are averaged to create an ensemble model [25]. The accuracy of the ensemble model, along with a confusion matrix, is computed, and a classification report is generated [26, 27].

IV. Results and Discussion

The results section presents the findings and performance metrics of the sentiment analysis conducted on the COVID-19-related Twitter dataset. This analysis utilized a comprehensive methodology that included data preprocessing, machine learning, deep learning, and an ensemble model to categorize tweets as "Negative," "Neutral," or "Positive" [28, 29]. The assessment metrics provide insight into the effectiveness of each method employed, facilitating a comparative analysis of model performance to determine the most suitable approach for this dataset. The fine-tuned BERT-based model demonstrated significant efficacy, achieving a test accuracy of 0.884. The confusion matrix for the BERT model is illustrated in Fig. 9, while a detailed classification report in Fig. 10 offers a thorough evaluation of its classification capabilities. In contrast, the CNN model designed for sentiment analysis achieved a commendable test accuracy of 0.817. The model's performance is further elucidated through visual representations, including a confusion matrix depicted in Fig. 11 and a detailed classification report in Fig. 12. Notably, the ensemble model, which aggregates predictions from both the CNN and BERT models, attained a test accuracy of 0.90. Fig. 13 showcases the ensemble model's effectiveness in accurately categorizing tweets into their respective sentiment categories, supplemented by a confusion matrix. The classification report for the ensemble model is presented in Fig. 14, providing additional insights into its performance. Overall, these results underscore the effectiveness of the applied methodologies in sentiment classification tasks.

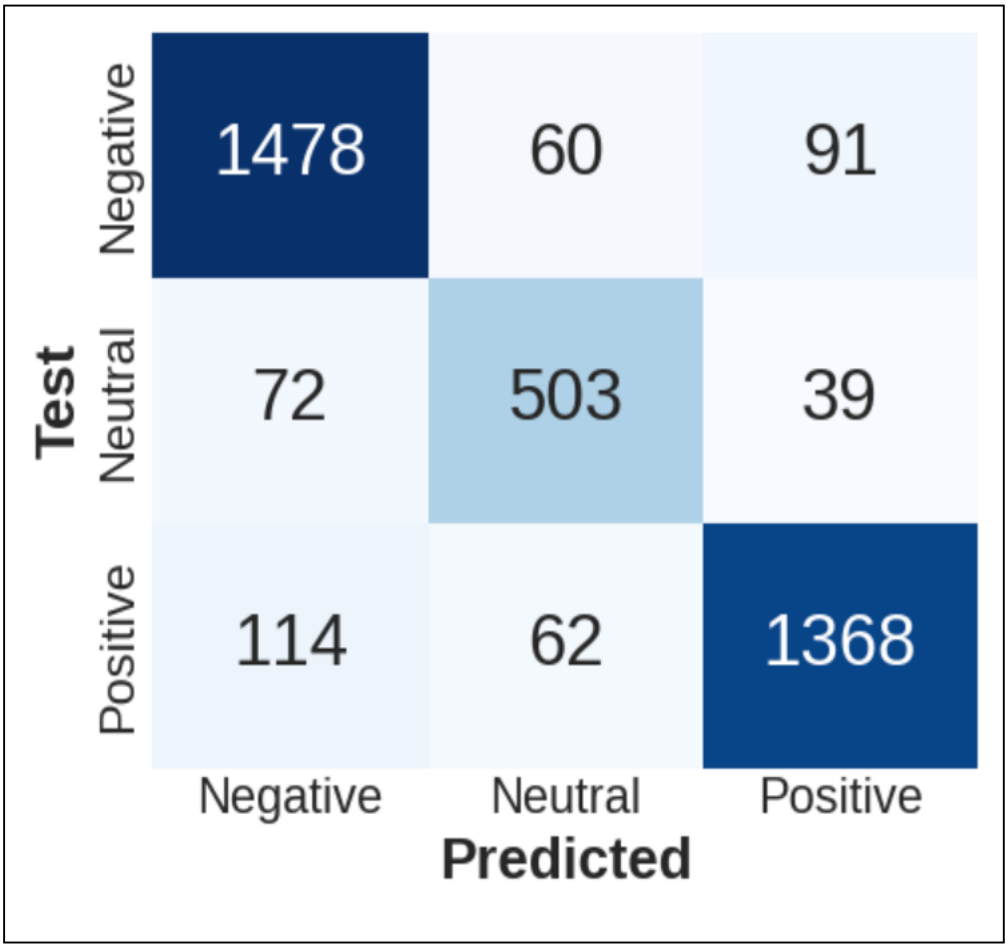


Fig. 9. Confusion matrix of BERT sentiment analysis.

Classification Report for BERT:

	precision	recall	f1-score	support
Negative	0.89	0.91	0.90	1629
Neutral	0.80	0.82	0.81	614
Positive	0.91	0.89	0.90	1544
micro avg	0.88	0.88	0.88	3787
macro avg	0.87	0.87	0.87	3787
weighted avg	0.88	0.88	0.88	3787
samples avg	0.88	0.88	0.88	3787

Fig. 10. Classification report for BERT.

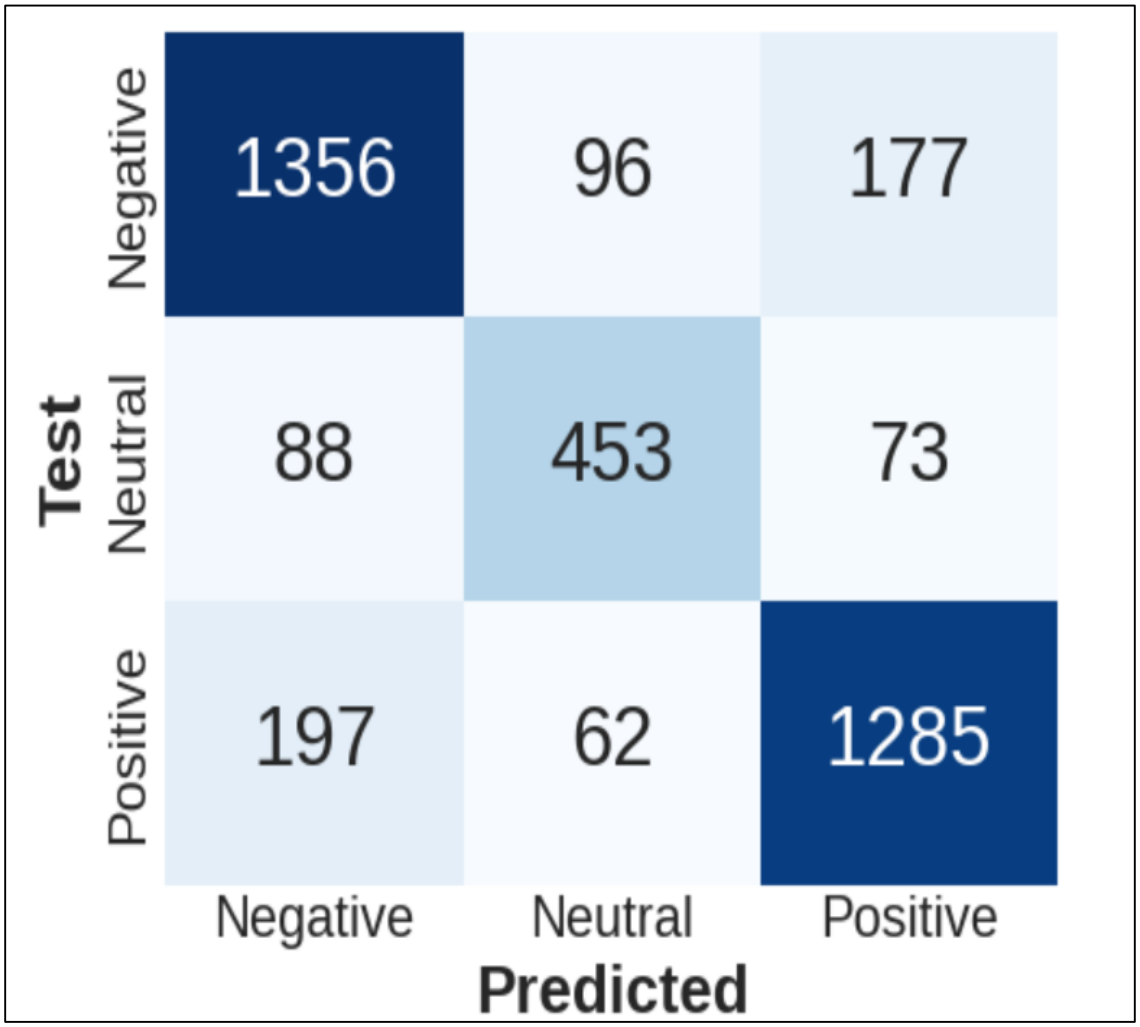


Fig. 11. CNN confusion matrix.

CNN Classification Report:

	precision	recall	f1-score	support
Negative	0.83	0.83	0.83	1629
Neutral	0.74	0.74	0.74	614
Positive	0.84	0.83	0.83	1544
accuracy			0.82	3787
macro avg	0.80	0.80	0.80	3787
weighted avg	0.82	0.82	0.82	3787

Fig. 12. CNN classification report.

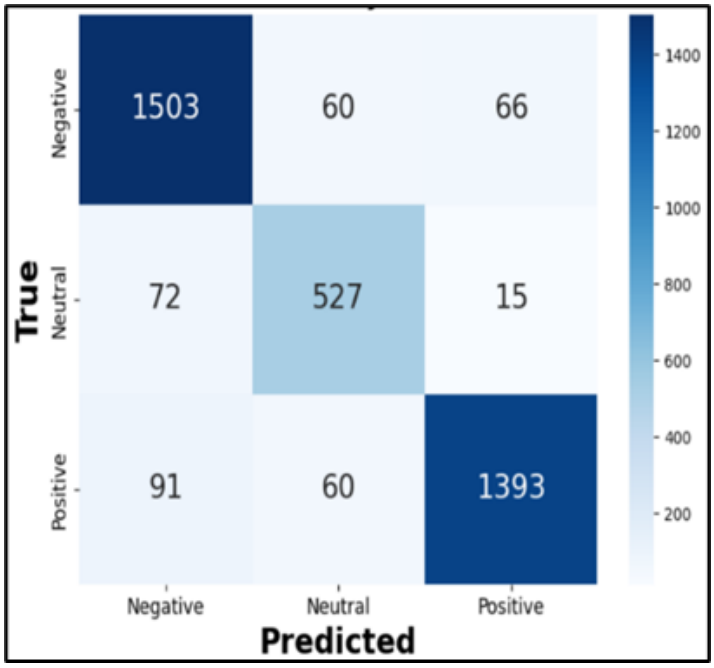


Fig. 13. Confusion matrix of ensemble model.

Classification Report:

	precision	recall	f1-score	support
Negative	0.90	0.92	0.91	1629
Neutral	0.81	0.86	0.84	614
Positive	0.95	0.90	0.92	1544
accuracy			0.90	3787
macro avg	0.89	0.89	0.89	3787
weighted avg	0.91	0.90	0.90	3787

Fig. 14. Ensemble model classification report.

In cases where the BERT model outperforms the ensemble, the results reflect the BERT model's dominant influence within the ensemble framework. This reliance underscores the necessity for a more sophisticated approach to balance the contributions of different models. Simply averaging predicted probabilities may not sufficiently mitigate the impact of one model over another, especially when there are significant imbalances in prediction probabilities. Despite this challenge, the ensemble model remains the preferred choice due to its robustness, stability, and versatility in handling diverse data patterns. This study focused on sentiment analysis of COVID-19-related tweets, providing practical insights for various sectors. Stakeholders in social sciences, digital marketing, public health, and policymaking can leverage these insights to understand public sentiment effectively. Sentiment analysis plays a critical role for public health officials, policymakers, social scientists, marketers, and emergency responders by tracking public opinion during crises. It informs decision-making, enables strategic adjustments, and enhances communication efforts, ensuring that responses align with public sentiment during emergencies. A comparative analysis with existing research, is presented in Table I.

TABLE I. COMPARATIVE STUDY WITH EXISTING LITERATURE

Ref.	Multiple ML models were tested before the final selection	Ensemble model selected	Data Preparation	Data Balancing	Data Vectorization	Results		
						F1 Score	Recall	Precision
[30]	NO	Yes	Yes	NO	Yes	High	High	High
[31]	Yes	NO	Yes	NO	NO	High	High	Satisfactory
[32]	NA	Yes	Yes	NO	Yes	High	High	High
[33]	Yes	No	Yes	No	No	NA	NA	High
Proposed work	Yes	Yes BERT and CNN	Yes 4 stage Data Preparation	Yes, using Random Oversampling	Yes, using the Count Vectorizer and TF-IDF transformer	High	High	High

V. CONCLUSION

In our study, a range of methods, from advanced deep learning techniques to conventional approaches, were discussed and assessed for the construction and assessment of sentiment analysis models. The work illustrated the revolutionary benefits of BERT, a cutting-edge transformer-based sentiment analysis methodology. About the test dataset, the BERT-based model demonstrated an impressive accuracy of 88.4%. The Naive Bayes model which had a respectable accuracy of 70% despite its streamlined architecture was also investigated. A CNN model was also a part of the investigation, and it demonstrated a commendable accuracy of 83%. Finally, an ensemble technique that merged predictions from BERT and CNN showed an accuracy of 89%, highlighting the potential of ensemble approaches to improve overall performance. Overall, the findings substantiate a combination of BERT and ensemble learning for enhancing sentiment analysis research across various domains with reliable and accurate classification. This study scouts the subtleties of the COVID-19 tweet sentiment analysis as well as the usefulness of sentiment analysis in a variety of contexts. Using state-of-the-art NLP and ensemble models of ML, the study offers an asset for anyone aiming to use public sentiment data for informed decision-making and for enhanced communication. For future research, expanding the sample size would enhance the evaluation of model performance. The results can also be utilized for policy-making in certain domains.

REFERENCES

[1] Rustam, F., Khalid, M., Aslam, W., Rupapara, V., Mehmood, A., Choi, G.S. (2021). A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. Plos One, 16(2): e0245909. https://doi.org/10.1371/journal.pone.0245909

[2] Mansoor, M., Gurumurthy, K., Prasad, V.R. (2020). Global sentiment analysis of COVID-19 tweets over time. arXiv preprint arXiv:2010.14234. https://doi.org/10.48550/arXiv.2010.14234

[3] Ayman Samir, Saleh Mesbah Elkaffas and Magda M. Madbouly(2021). X (formerly Called Twitter) Sentiment Analysis using BERT(2021). ICCTA 2021, 11-13 December, Alexandria, Egypt

[4] Zou, H., Wang, Z. A semi-supervised short text sentiment classification method based on improved Bert model from unlabelled data. J Big Data 10, 35 (2023). https://doi.org/10.1186/s40537-023-00710-x

[5] Liao, S., Wang, J., Yu, R., Sato, K., Cheng, Z. (2017). CNN for situations understanding based on sentiment analysis of twitter data. Procedia Computer Science, 111: 376-381. https://doi.org/10.1016/j.procs.2017.06.037

[6] Saleena, N. (2018). An ensemble classification system for twitter sentiment analysis. Procedia Computer Science, 132: 937-946. https://doi.org/10.1016/j.procs.2018.05.109

[7] Hussein, D.M.E.D.M. (2018). A survey on sentiment analysis challenges. Journal of King Saud University-Engineering Sciences, 30(4): 330-338. https://doi.org/10.1016/j.jksues.2016.04.002

[8] Wankhade, M., Rao, A.C.S., Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55(7): 5731-5780. https://doi.org/10.1007/s10462-022-10144-1

[9] Villavicencio, C., Macrohon, J.J., Inbaraj, X.A., Jeng, J. H., Hsieh, J.G. (2021). Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes. Information, 12(5): 204. https://doi.org/10.3390/info12050204

[10] Joloudari, J.H., Hussain, S., Nematollahi, M.A., et al. (2023). BERT-deep CNN: State of the art for sentiment analysis of COVID-19 tweets. Social Network Analysis and Mining, 13(1): 99. https://doi.org/10.1007/s13278-023-01102-y

[11] Chintalapudi, N., Battineni, G., Amenta, F. (2021). Sentimental analysis of COVID-19 tweets using deep learning models. Infectious Disease Reports, 13(2): 329-339. https://doi.org/10.3390/idr13020032

[12] Liu, Yi & Lu, Jiahuan & Yang, Jie & Mao, Feng. (2020). Sentiment analysis for e-commerce product reviews by deep learning model of Bert-BiGRU-Softmax. Mathematical Biosciences and Engineering. 17. 7819-7837. 10.3934/mbe.2020398.

[13] Kaur, H., Ahsaan, S.U., Alankar, B., Chang, V. (2021). A proposed sentiment analysis deep learning algorithm for analyzing COVID-19 tweets. Information Systems Frontiers, 23(6): 1417-1429. https://doi.org/10.1007/s10796-021-10135-7

[14] Chakraborty, K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R., Hassanien, A. E. (2020). Sentiment analysis of covid-19 tweets by deep learning classifiers—A study to show how popularity is affecting accuracy in social media. Applied Soft Computing, 97: 106754. https://doi.org/10.1016/j.asoc.2020.106754

[15] Unlu, A., Truong, S., Tammi, T. and Lohiniva, A.L., 2023. Exploring political mistrust in pandemic risk communication: mixed-method study using social media data analysis. Journal of Medical Internet Research, 25, p.e50199. doi:10.2196/50199

[16] Jain, R., Bawa, S., Sharma, S. (2022). Sentiment analysis of COVID-19 tweets by machine learning and deep learning classifiers. In Advances in Data and Information Sciences: Proceedings of ICDIS 2021, pp. 329-339. https://doi.org/10.1007/978-981-16-5689-7_29

[17] Braig, N., Benz, A., Voth, S., Breitenbach, J., & Buettner, R. (2023). Machine learning techniques for sentiment analysis of COVID-19-related twitter data. IEEE Access, 11: 14778-14803. https://doi.org/10.1109/ACCESS.2023.3242234

[18] Pipal, C., Schoonvelde, M., Schumacher, G., & Boiten, M. (2024). JST and rJST: joint estimation of sentiment and topics in textual data using a semi-supervised approach. Communication Methods and Measures, 1–19. https://doi.org/10.1080/19312458.2024.2383453

[19] Acheampong, F.A., Nunoo-Mensah, H. & Chen, W. Transformer models for text-based emotion detection: a review of BERT-based approaches.

Artif Intell Rev 54, 5789–5829 (2021). https://doi.org/10.1007/s10462-021-09958-2

[20] A. A. Maruf, F. Khanam, M. M. Haque, Z. M. Jiyad, M. F. Mridha and Z. Aung, "Challenges and Opportunities of Text-Based Emotion Detection: A Survey," in IEEE Access, vol. 12, pp. 18416-18450, 2024, doi: 10.1109/ACCESS.2024.3356357.

[21] Drus, Z. and Khalid, H., 2019. Sentiment analysis in social media and its application: Systematic literature review. Procedia Computer Science, 161, pp.707-714.

[22] Wang, S., Huang, M. and Deng, Z., 2018, July. Densely connected CNN with multi-scale feature attention for text classification. In IJCAI (Vol. 18, pp. 4468-4474). https://dl.acm.org/doi/abs/10.5555/3304222.3304391

[23] Wangpoonsarp, Attaporn & Shimura, Kazuya & Fukumoto, Fumiyo. (2020). Unsupervised Predominant Sense Detection and Its Application to Text Classification. Applied Sciences. 10. 6052. 10.3390/app10176052.

[24] Kazmaier, J. and Van Vuuren, J.H., 2022. The power of ensemble learning in sentiment analysis. Expert Systems with Applications, 187, p.115819. https://doi.org/10.1016/j.eswa.2021.115819

[25] Minaee, S., Azimi, E. and Abdolrashidi, A., 2019. Deep-sentiment: Sentiment analysis using ensemble of cnn and bi-lstm models. arXiv preprint arXiv:1904.04206.

[26] Kausar, M.A., Soosaimanickam, A. and Nasar, M., 2021. Public sentiment analysis on Twitter data during COVID-19 outbreak. International Journal of Advanced Computer Science and Applications, 12(2).

[27] Agrawal, R., Singh, J., Ghosh, S.M. (2020). Performance Appraisal of an Educational Institute Using Data Mining Techniques. Computing in Engineering and Technology. Advances in Intelligent Systems and Computing, vol 1025. Springer, Singapore. https://doi.org/10.1007/978-981-32-9515-5_69

[28] Y U. Naseem, I. Razzak, M. Khushi, P. W. Eklund and J. Kim, "COVIDSenti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis," in IEEE Transactions on Computational Social Systems, vol. 8, no. 4, pp. 1003-1015, Aug. 2021, doi: 10.1109/TCSS.2021.3051189.

[29] Zhao, A. and Yu, Y., 2021. Knowledge-enabled BERT for aspect-based sentiment analysis. Knowledge-Based Systems, 227, p.107220. https://doi.org/10.1016/j.knosys.2021.107220

[30] Sunitha, D., Patra, R.K., Babu, N.V., Suresh, A., Gupta, S. C. (2022). Twitter sentiment analysis using ensemble based deep learning model towards COVID-19 in India and European countries. Pattern Recognition Letters, 158: 164-170. https://doi.org/10.1016/j.patrec.2022.04.027

[31] Xie, Tianyi & Ge, Yaorong & Xu, Qian & Chen, Shi. (2023). Public Awareness and Sentiment Analysis of COVID-Related Discussions Using BERT-Based Infoveillance. AI. 4. 333-347. 10.3390/ai4010016

[32] Rustam, F., Khalid, M., Aslam, W., Rupapara, V., Mehmood, A. and Choi, G.S., 2021. A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. Plos one, 16(2), p.e0245909

[33] Alabid, N.N. and Katheeth, Z.D., 2021. Sentiment analysis of twitter posts related to the covid-19 vaccines. Indonesian Journal of Electrical Engineering and Computer Science, 24(3), pp.1727-1734. DOI: 10.11591/ijeecs.v24.i3.

Visual Translation of Auspicious Beliefs in Quanzhou Xi Culture from the Perspective of Man-Machine Collaboration

Li Zheng*[1], Xu Zhang[2], Huiling Guo[3]
College of Software Engineering, Xiamen University of Technology, Xiamen, China[1, 2]
International College of Xiamen University, Xiamen University, Zhangzhou, China[3]

***Abstract*—The "Xi" concept in the inheritance of auspicious culture covers the abundance of spiritual and material life, its symbolism is gorgeous and timeless and has lasted for thousands of years. Objective: This study investigates the Quanzhou "happiness" culture, which embodies "reverence for virtue and auspicious beliefs," exploring its visual symbolization, graphical derivation, redesign, and innovative cultural expressions. Methods: Utilizing literature analysis, field research, and a combination of shape grammar and artificial intelligence, this study dissects and evolves the visual symbols of Quanzhou Xi culture to achieve innovative design through human-machine collaboration. Results: The study deeply refines representative visual symbols of Quanzhou happiness culture, including the "卍" character from Quanzhou embroidery, the Eight Immortals color for wedding happiness, and the longevity turtle cake stamp for longevity happiness. It analyzes and demonstrates the innovative practice of these visual symbols, establishes a folklore perspective, and transitions the happiness culture into a modern fashion context. Conclusion: The research constructs a visual symbol folklore perspective of Quanzhou Xi culture, providing a systematic theoretical foundation and innovative practice paths for promoting and inheriting Xi culture in the modern design field. It promotes Quanzhou Xi culture's innovative application and fashion transformation in contemporary design.**

***Keywords*—Quanzhou Xi culture; symbol visualisation; shape grammar; artificial intelligence; man-machine collaborative design**

I. INTRODUCTION

'Xi' means joy and happiness. From symbol auspiciousness to visual symbolization, both convey feelings of satisfaction, good fortune, and happiness. "Long drought meets sweet rain, foreign lands meet old acquaintances, bridal chamber flowers and candles night, and when one achieves success in the imperial examination" [1]. The symbolic meaning of "Xi" today is the visual communication method of images. Since the initial definition of "Xi" in the Song Dynasty, it refers to the folk celebration of happy events, mostly referring to men's and women's marriages, childbirth, and so on. Quanzhou, with a thousand-year history of navigation and exchanges between Chinese culture and overseas cultures, is known for its unique regional culture that embodies the auspicious fusion of the Song and Yuan dynasties "Individuals hailing from ten states" and "seeking common ground while reserving differences".

Visual symbolization refers to the practical representation of images conveyed through symbols, and the design expression of cultural symbols can be comprehended by recreating or reorganizing cultural symbols. Liao et al. proposed to interpret and assess the meaning of visual symbols and organically integrate the visual reality and objective reality of cultural symbols with natural meaning, normative meaning, and intrinsic meaning [2]. Li contends that China is constrained by the Western modernist design concept that merely emphasizes function and technology, resulting in the deficiency of its own history and culture. Chinese design demands its own language, and to restore these cultures, design support is necessary [3]. Culture bestows a soul upon design, and design showcases the vitality of culture. Outstanding design is never rootless water; it must be deeply rooted in traditional national culture and continuously evolve into modern aesthetics through cultural integration and design innovation. As proposed by Xia, the visual translation of the artistic images presented by the pleasant culture of Quanzhou is a significant aesthetic activity in human spiritual life and a means to understand the objective world through artistic images [4]. In the application of visual symbol innovation, Xi culture mainly focuses on cultural and creative products, brand design, packaging design application, and exploration of cultural connotations. Mo et al. investigated the connotation of traditional Xi culture and the form elements that constitute the "Xi" schema, deconstructed and reconstructed the "Xi" schema, and passed on the practical value and cultural value of Xi culture with cultural and creative products as the carrier [5]. Yuan, through his research on the folk culture aesthetics of Quanzhou's "Pu" women's headwear, put forward the philosophical thought of "harmony between nature and man", which coexists and integrates with the local culture [6]. There is still a scarcity of a complete research system and an accumulation of achievements in the analysis and study of the symbols of Xi culture in Quanzhou.

This research aims to clarify the characteristics of Quanzhou's "Xi" culture and analyze its connotations, conducting a study on the visual symbolization of Quanzhou's "Xi" culture, and exploring its value in the fields of art and culture through human-computer collaborative design, thereby reshaping the artistic expression of Quanzhou's "Xi" culture. The main innovative contributions of this study are reflected in the following three points: (1) A deep refinement of Quanzhou's "Xi" culture visual symbols and the construction of a folk perspective; (2) The use of human-computer collaboration to further develop and redesign the visual symbols of Quanzhou's "Xi" culture, achieving an organic integration of art and

technology, creativity, and algorithms; and (3) The reconstruction of the modern and cultural value of Quanzhou's "Xi" culture visual symbols, providing new pathways for the innovative application and fashion-oriented transformation of "Xi" culture in modern design fields.

II. Return To Faith: Quanzhou Xi Culture To be Unfolding

Quanzhou is renowned as "Individuals hailing from ten states" and "the forerunner of the Maritime Silk Road", and the custom of "Worship Mude" continuously gave rise to the early marriage and birthday celebration culture. Subsequently, due to the influence of a significant number of overseas Chinese returning to their hometowns throughout history, they brought back the auspicious culture of Southeast Asia and the belief culture of southern Fujian, and gradually integrated, nurtured, and developed the distinctive Xi culture of Quanzhou in the collision and accumulation of various cultures. "Tea plate round sweet tea sweet, two family marriage at the end of winter born twins, Children, and grandchildren, all being eager to learn" is a popular song about the Xi culture in the Quanzhou region. It demonstrates the inheritance and promotion of traditional values such as family concept, family affection, marriage, and descendants, and possesses regional style, converges diverse cultures, undergoes inclusive development with profound cultural forms, and serves as a bond of cultural identity and emotional exchange in the Quanzhou area. As Geertz explained, the analysis of culture is not an experimental science in search of rules, but an interpretive science in search of meaning [7].

A. The Enduring Xi Culture of Quanzhou

The Xi culture of Quanzhou is reflected in daily life which is to worship Mude, which is rooted in the tradition of Mindi, and has the connotation of cultural integration and renewal. "Wenling Old Story" describes "worship MuDe" culture as "joy", recorded in the competition to choose a day to worship the Lord of heaven and pray for blessing. The visual symbols of birthday, marriage, birth, and so on are directly evolved from the tradition, while many modern representations of Xi are produced in the unique style of Quanzhou's overseas Chinese culture and traditional culture. This so-called culture is not firmly "sealed" in the tradition, but refined and excellent integration outside the region. The perfect fusion of tradition and modernity has promoted Quanzhou Xi culture to be widely inherited and continuously developed in the local area, and deeply integrated into the folk culture, which has a deep and huge influence on contemporary society. The continuation and promotion of Quanzhou Xi culture can not only help the people feel the culture and history of their hometown, and get close to the wisdom and wealth of their ancestors, but also become more orderly, interesting, tasteful, and homelike in their daily life. The core value of Mindi traditional culture has cultivated Quanzhou Xi culture, and the Quanzhou Xi culture with the continuity of "worship Mude" has made Mindi culture more colorful.

B. A Hundred Flowers Blossom of QuanZhou Xi Culture

Quanzhou Xi culture takes Minnan belief as its source and Nanyang auspiciousness as its stream. It presents diverse and unique visual symbols, which integrate the inheritance of inland civilization and marine culture. "Have faith in ghosts and sorcery", especially among people in southern Fujian. The people of southern Fujian "respect deities as if He were present" and "prefer to believe in His existence rather than His absence". The aesthetic features of Quanzhou Xi culture's visual symbols are in accordance with the aesthetic orientation of sacrificial ceremonies, taking southern Fujian faith as the mainstream, close to the folk, deeply rooted in life, rich in the grassroots local flavor, human touch, and always maintain its vigorous vitality. The opening and expansion of the Maritime Silk Road brought together people from ten prefectures and auspicious cultures brought back from overseas Chinese villages, etc., which promoted Quanzhou to better integrate the core Chinese culture, other regional cultures, and the essential elements of foreign cultures in its internal cultural structure, thereby deepening its own cultural characteristics [8]. Influenced by the Southeast Asian and Western cultures of overseas Chinese, Quanzhou Xi culture has given rise to diversified integration and development. For example, Quanzhou wedding customs, are affected by foreign wedding gowns, since Quanzhou wedding customs do not wear white, the people will change the wedding dress to red and transform it into a Qipao. The unique blending and collision of Chinese and Western cultures present a multi-integrated development form of Quanzhou Xi culture, not only maintaining the variation of South Asian auspicious culture but also adhering to its core values in the pursuit of the core mainstream culture of southern Fujian faith, and developing in the process of modernization.

C. Formation and development of Quanzhou Xi culture

Folk marriage is referred to as "great joy", a women's pregnancy is known as "you have joy", children are termed "Add Xi", and Eggs to celebrate the birth of a baby are called "happy eggs". Marriage is joyous and having children is also a joy [9]. In Quanzhou and even across the entire country, since the advent of new life, it is all about "Xi". The first cry of a newborn is a "Xi" sound, and the first year is also a joyous event. When one reaches the age of 16, they become an adult, encounter a lover, and enter into marriage. There is also the joy of having a child, birthdays, etc. The "Five Xi culture" comes from the real scene exhibition area of the "Life Five Xi Wall" in the Quanzhou Marriage Festival Folk Culture Museum, presenting the most significant joys in the lives of Quanzhou people, namely the joy of birth, the joy of the first year, the joy of adulthood, marriage, and birthdays. Table I is a classification and analysis of the content presentation and existing development of the five great happiness cultures in Quanzhou.

TABLE I. CLASSIFICATION OF THE COMPONENTS OF QUANZHOU XI CULTURE

Compositional classification	Joy of birth	The joy of grasping week	The joy of adulthood	Wedding joy	Make a birthday celebration
Content rendering	Newborn clothing embroidery "Wan" word, package "Floral quilt", eat red egg, full moon feast	Clean hands, roll disaster, crown, comb hair, grasp week, pick handprints, eat fu, seal wine, Family photo	Prepare the "three sacrifices", go to the temple to pay deities set a banquet, grandma's house As usual to send clothes, hats, shoes and socks, eggs and so on	Blind date, home visit, talk about betrothal gifts and clothes and pants money, tie Red envelopes, Lift plates, on top, tie spring flowers	Set up longevity hall, point longevity candle, eat Longevity noodles, set a birthday banquet, do Turtle kueh
Existing development	Traditional handicraft and traditional culture continue, but the scope of continuation gradually narrowed	Some traditional customs and cultural forms have gradually simplified or even disappeared, and the industry Chain related to Zhuazhou object has developed rapidly	The process is gradually simplified and the form is more modern	Traditional customs remain intact, and some bad habits such as high endowment have been abandoned, but traditional customs in urbanized areas have been simplified and gradually Westernized	Traditional ways And scope are decreasing, and Traditional gifts such as longevit Y noodles, longevity kueh, longevity peaches, longevity hangingpictures, etc, have been replaced by more high-end and customized products

III. BE CONTENT WITH ONE'S HEART: THE CONNOTATION EXPRESSION OF VISUAL SYMBOLS

The visual symbol serves as a means of expression in design and is collectively known as a visual symbol. It represents the symbolic reconstruction or reorganization of traditional images or abstract concepts, utilizing symbols such as line, light, shadow, color, intensity, expression, balance, and form to showcase modern design ideas and connotations, convey information, and exhibit its beauty and internal equilibrium. With its distinctive continuity and diversity, the visual symbols of Quanzhou Xi culture prominently demonstrate the emotional resonance of auspicious meaning and people's modern aesthetic sense. The visual symbols of early Quanzhou Xi culture primarily centered on image art, like painting, sculpture, architecture, etc. In the modern context, visual symbols transcend the realm of image art itself and prevalently emerge in auspicious belief forms such as folk custom, sacrifice, and birthday celebrations, playing a vital role in the visual communication of clothing, decoration, and adornment in modern life and constituting an indispensable part of modern cultural symbols. Due to its outstanding design display and cultural expression, it possesses a visual form accessible to human eyes and ears on the material form and an enlightening situation at the spiritual level.

The visual symbols of Quanzhou Xi culture are utilized in the domains of clothing, decoration, and decorative goods, and can be categorized into three types in accordance with their expression content and function, as depicted in Table II. The first type is the indicative symbol of Xi culture, such as the Quanzhou embroidered "囍" character, which conveys the happiness of life in a simple and explicit symbol form to achieve efficient dissemination of information and prompt understanding. The second category is the cultural image symbols, such as the wedding of the typical representative of Quanzhou's eight immortal colors, through the employment of images, graphics, icons, and other visual elements, to guide, stimulate, and attract the audience's attention, emotions, and thinking, and convey various kinds of information, appeals, and cultural symbols. The third type is the symbolic symbol of Xi culture, which is associated with a certain fixed symbol and a specific meaning, representing a specific concept or symbolizing a profound meaning, such as the turtle pattern being regarded as a symbol of longevity, and green leaves representing life and nature, etc. This symbolization can assist in conveying and reinforcing a certain emotion or value, enabling the audience to have a profound resonance and identification.

TABLE II. CLASSIFICATION OF VISUAL SYMBOLS IN QUANZHOU XI CULTURE

Constituent classification	Indicative symbol	Pictorial symbol	Symbolic symbol
Symbolic feature	Strong directness, efficient transmission, rapid understanding	High eye attraction, easy to trigger emotional resonance	Symbols are associated with meanings and have deep representation meanings
Quanzhou Xi cultural symbol	Quanzhou embroidery version "囍" character, seven niang mother, etcs	Eight immortals color, paper cutting, etcs	Longevity turtle cake stamp, Xiniang, Zha spring flower, etcs

Besides the aforementioned three types of symbols, visual symbols can also be symbolized in accordance with visual elements like color, text, texture, etc., to maximize and effectively extract information during the process of expressing visual information and meaning.

IV. MAN-MACHINE COLLABORATION: THE VISUAL SYMBOL OF QUANZHOU XI CULTURE

With the rapid advancement of artificial intelligence technology, its introduction has ushered in a novel paradigm to

the domain of visual design and profoundly facilitated the practice of human-machine collaboration. Under this novel paradigm, artificial intelligence and designers can undertake the same or different design tasks through specific information exchange and cooperation mechanisms, collaborate to accomplish the predefined design goals, and bring about the innovation and breakthrough of visual design. In the process of visual design, AI is relatively potent in competitive and predictable workflows, and is particularly adept at handling complexity and multitasking; Human intelligence is flexible and creative during the design process, and excels in knowledge comprehension and strategic thinking, as depicted in Fig. 1 [10]. The application of human-machine collaborative design pertains to the processing of Media collaborative technology.

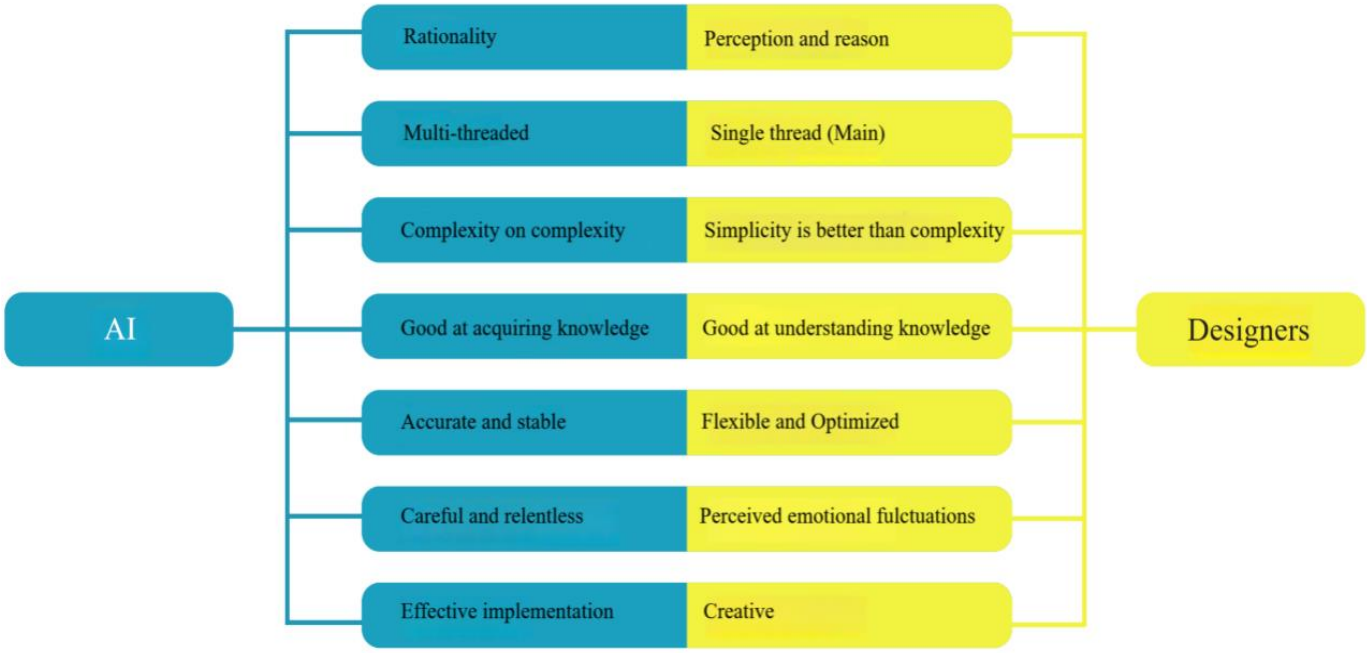


Fig. 1. Comparative paradigms of human-machine collaboration in visual symbol design.

A. Visual Symbols Driven by Shape Grammar

Shape grammar is a production system that generates a specific category of geometric shapes. It comprises shape rules and a generation engine for the selection and processing of rules. Originally put forward by George Stiny and James Gips in 1972 based on the theoretical study of "symbolic language", they applied the initial shape rules to generate new geometric shapes and style design and analysis approaches [11].

Generally, shape grammar adheres to aesthetic principles and geometric composition, and employs rotation, movement, arrangement, scaling, displacement, and mirror switching to infer specific or extracted shapes. In the actual design and application process, there is a considerable amount of intersection and repetition of shape inference rules. As depicted in Fig. 2, during the generation of shape grammar graphics, the initial specific shape is selected, the inference rule base is chosen, the input constraint size is specified for morphological reasoning, and the graphics are organized, screened, and reorganized in combination with modern design languages to obtain new visual symbols.

With the rapid development of artificial intelligence technology, the integration of shape grammar and AI has brought unprecedented innovative opportunities to the field of visual symbol design. AI algorithms can automatically learn and optimize shape inference rules, improving generation efficiency and accuracy. Additionally, AI can dynamically adjust rule sets and input constraints based on user feedback and design requirements, achieving personalized and intelligent design. In terms of the specific integration path, this paper will use the cases of the "Quanzhou embroidered 'Wan (卍)'," "Eight Immortals Color," and "Longevity Turtle Cake Stamp" to illustrate the methods for analyzing and redesigning visual symbols.

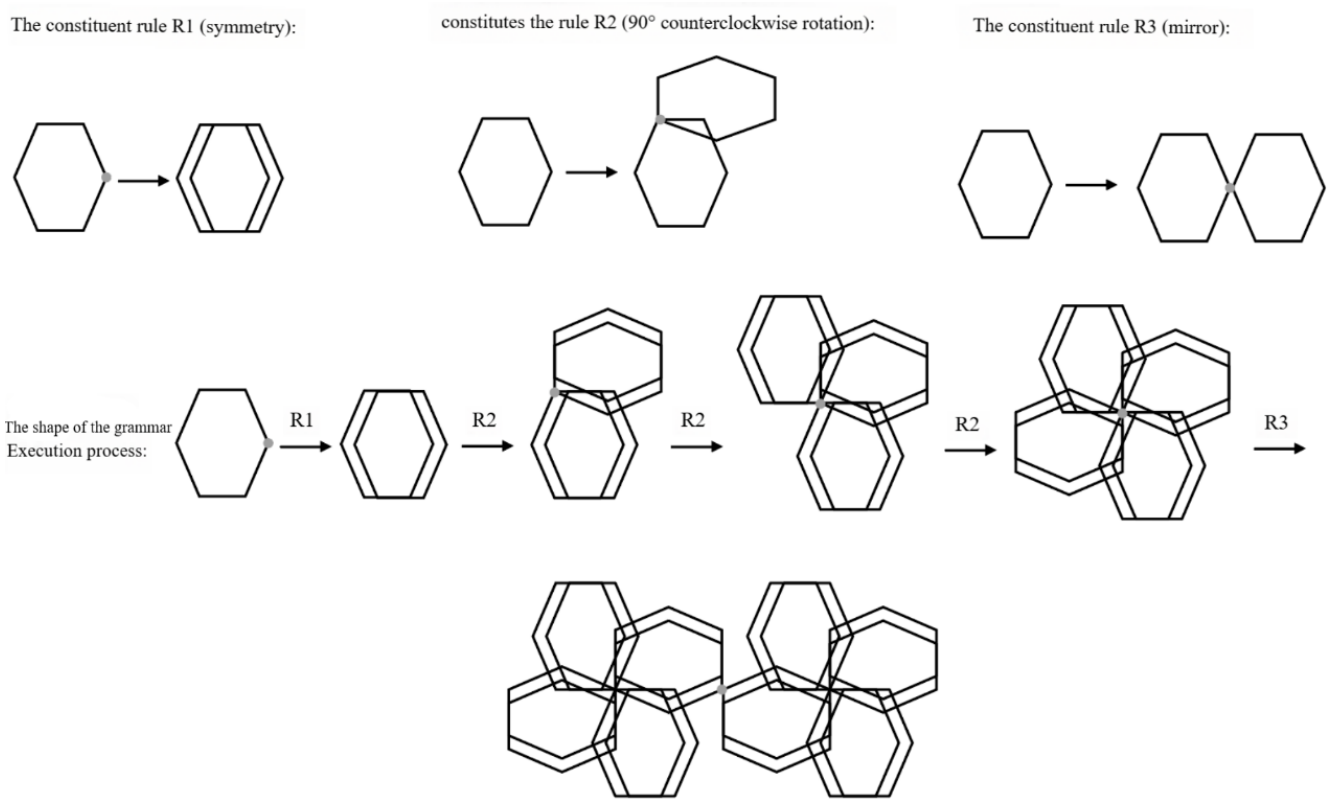


Fig. 2. Shape grammar graphic generation process.

B. Visual Symbols for Artificial Intelligence Operations

The application of artificial intelligence in the domain of graphic design constitutes an inevitable tendency. Sora, OpenAI's most recent novelty, is an AI model for generating video content based on textual prompts. The crux of Sora lies in its capacity to comprehend and execute complex visual renderings, construct intuitive physics, and even undertake long-term reasoning and semantic foundations. This implies that designers can merely furnish a one-sentence description and Sora is capable of generating a realistic video, which holds vast potential for brand storytelling, advertising creativity, product presentation, and beyond. The popularity of OpenAI constitutes a manifestation of the substantial influence exerted by the advancement of artificial intelligence technology on human production and life [12].

In March 2022, the US Research Laboratory released the first artificial intelligence painting program named Midjourney, which applies artificial intelligence technologies such as deep learning, image recognition, and natural language processing to analyze the text provided by users and generate visual graphics related to it. The program can recognize the distinctive artistic styles of different artists, and learn and analyze a wide range of works of art and design patterns to generate unlimited design suggestions that stimulate the creativity of designers and enable them to explore and realize their design concepts more efficiently. When this paper discusses the application of human-machine collaborative design in visual design, Midjourney, as a powerful AI image generation tool, offers a wealth of cases and revelations.

C. Man-machine Collaboration Design Application of Quanzhou Xi Cultural Visual Symbols

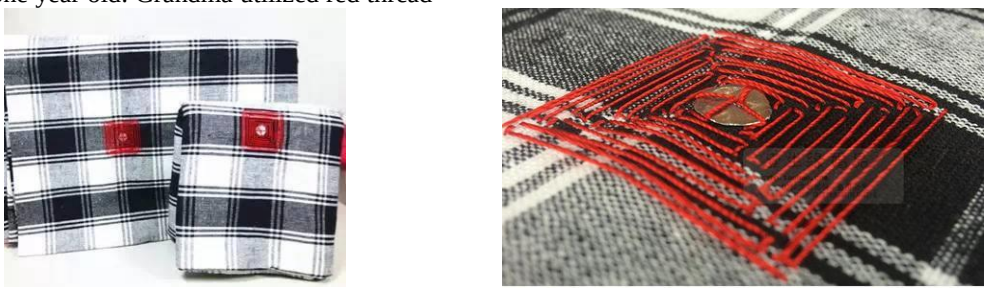
The collaboration between human intelligence and artificial intelligence varies across diverse fields. Human intelligence takes the lead in tasks that demand creativity, strategy, and empathy. In contrast to the "visual average" of the images produced by artificial intelligence, human intelligence possesses a distinctive emotional touch and understanding of culture. Artificial intelligence offers designers abundant inspiration and creativity through in-depth analysis of a wide range of design data and trends. It also significantly broadens the boundaries of design. The two complement each other, co-create symbiotically, and innovate brilliantly.

Widely utilized in Quanzhou Xi cultural visual symbols mainly consist of eight immortal colors, a pan, an embroidered version of the "Wan (卍)" character, turtle kueh printing, paper cutting, spring flowers, and Xi Niang. The symbol of happiness - Quanzhou embroidered "Wan (卍)" character is chosen, and the meaning of warding off evil and ensuring peace is conveyed through the red embroidered character; The image symbol of wedding joy - Eight immortal colors, expressing auspiciousness and adding happiness; The symbolic symbol of longevity - Longevity turtle cake stamp, symbolizing longevity and the meaning of protecting the house. The three representative symbols are subjected to graphic deduction and redesign through man-machine collaboration.

1) Indicative symbol - Quanzhou embroidered "Wan (卍)" character: The "卍" was originally a symbol of auspiciousness appearing on the chest of the Buddha. The clockwise rotating

wheel of the Buddha radiates light, symbolizing auspiciousness, blessings, and longevity. In the second year of Wu Zetian's longevity in the Tang Dynasty (693 AD), the "卍 " symbol was introduced into the Chinese character system, pronounced as "Wàn", meaning "a collection of auspicious and virtuous things"

According to the customs of Quanzhou, special ceremonies will be conducted when the newborn reaches one month old, four months old, and one year old. Grandma utilized red thread to embroider a pattern resembling a 卍 (as shown in Fig. 3) on new clothes or items such as black and white napkins as a ritual, and placed a small piece of lead on the embroidery, known as "lead money". The pronunciation of "lead" in the Quanzhou dialect is the same as "edge", meaning "There exists an intersection where someone loves", symbolizing grandma's love.



(a) (b)

Fig. 3. Quanzhou 'Huabei' embroidery pattern of the 卍 symbol.

The original 卍 embroidery technique was employed in Quanzhou in the early stage. This custom has been handed down to the present. Owing to its regularity and systematization, it has continuously evolved and innovated in two or four directions, as depicted in Fig. 4. The million-word pattern is also known as the million-word turn or the million-word brocade. The four ends and four outer extensions are interconnected, and folk people call it the million-word continuous head or continuous, to convey its meaning of wealth and continuity.

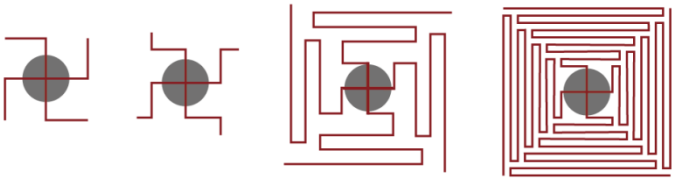


Fig. 4. Evolution of the Quanzhou embroidery pattern of the 卍 symbol.

Choose the 卍 of the square continuous and the nine-square continuous Quanzhou embroidery version, which symbolizes wealth good fortune, and double happiness Linmen, as the initial texture pattern, follows the shape rules such as full and full picture, combines the translation R1, rotation R2, and other deduction methods in the shape grammar to implement the graphic deduction, and take the most representative nine-square continuous Quanzhou embroidery version of the 卍 as the central extension point of Quanzhou Xi cultural graphics. A new graphic scheme is formed, and the design effect is presented as shown in Fig. 5. The new visual symbol of the redesign highlights the charm of the traditional symbol, integrates the modern design concept, and presents the change in the fashion image.

Fig. 5. The embroidery pattern of the 卍 symbol is designed with shape grammar.

Creative inspiration is the embroidered version of "卍" character graphics designed by shape grammar. Midjourney artificial intelligence drawing program is used to create graphic design works. Display the input text to generate a graphic effect as shown in Fig. 6. After adjusting the text many times, the author created a set of graphic renderings with the help of the software, but the software-generated graphic design works for art display is still good, and there is still a certain gap between the creation of Quanzhou Xi cultural heritage and humanistic feelings, and it is difficult to get a practical application in situations involving emotion, personality, and ethics.

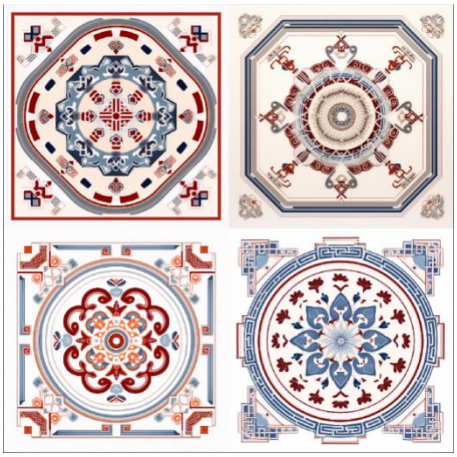


Fig. 6. Midjourney's design of the embroidery pattern of the 卍 symbol is designed with shape grammar.

Artificial intelligence explores diverse visual styles and expressions, rapidly generates a multiplicity of design schemes, and takes graphic creation as the structural framework. The designer extracts the Quanzhou embroidery version of the 卍 as the main decorative body, presenting a symmetrical or balanced form. The man-machine collaborative design (as shown in Fig. 7) breaks through the traditional design limitations imposed by human cognition and the capabilities of material calling and copy editing. However, artificial intelligence lacks the emotional expression of the character of the "卍" in Quanzhou embroidery. By giving full play to the advantages of artificial intelligence and human intelligence, exchanging their respective strengths, and complementing each other, a complete pattern and consistent style can be achieved. Simultaneously, it embodies the connotation of Quanzhou Xi culture.

Fig. 7. Embroidery pattern of the 卍 symbol was designed through human-machine collaboration.

2) Graphic Symbol - Eight Immortalis: Quanzhou folk marriage customs reveal that on the day before the marriage in the woman's home, a "red brocade" can be seen hanging on the beam. Locally, it is commonly known as "color". As depicted in Fig. 8, the red cloth, symbolizing happiness and excitement, is frequently used as the base, complemented by gold and silver threads, enhancing the noble atmosphere of "color". "Color" is adorned with various silk threads to depict the images of deities. Particularly, "Eight Immortals Crossing the Sea" is the most classic and widely spread one, commonly referred to as "Eight Immortals"

(a) (b)

Fig. 8. Quanzhou eight immortals color.

The Eight Immortals are derived from folklore. In the allusion, "Eight immortals" symbolizes the divine power of the people to pray for their wishes. They are regarded as a symbol to ward off evil spirits, implying good luck and adding joy. The rendering effect is shown in Fig. 9.

Fig. 9. Symbolic pattern of the eight immortal's color.

Based on the Eight Immortals symbol in Fig. 9, combined with the wave patternsymbolizing long-term prosperity and strength and the lotus pattern symbolizing purity andsteadfastness, and combined with the shape grammar of Rl composition rules (translation copy), R3 composition rules (mirror), etc., shape deduces and redesign, forming a regular rectangularshape, showing the symmetry rhythm beauty. The effect is shown in Fig. 10.

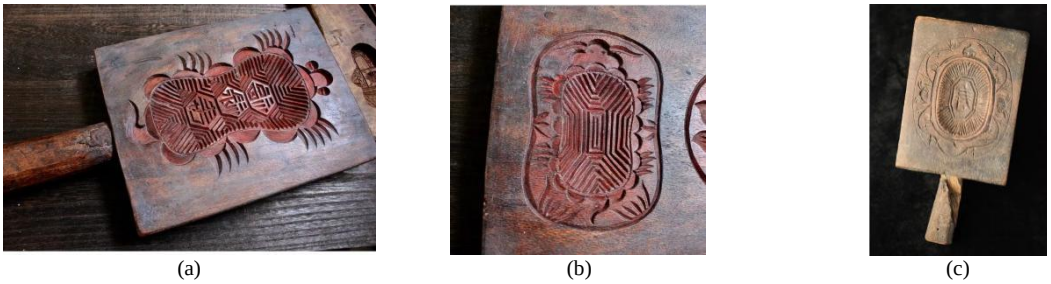
Referring to the eight Immanence symbol graphics designed by shape grammar, Midjourney establishes the creative inspiration source, and the generated effect is presented in Fig. 11. Learning from the generated structural paradigm, adhering to the color attribute of "bright red and precious yellow", and employing human-machine collaborative design. The design presentation effect is displayed in Fig. 12. Artificial intelligence can play the role of providing external stimuli in cooperation with human beings, triggering the designer's lateral thinking and influencing creative ideas and results. Artificial intelligence introduces foreign elements to provide human cognitive stimulation, and designers utilize individual consciousness to express the emotion of the "eight Immortals" will. The two cooperate complementarily to contribute content, accelerate the birth of ideas, and help overcome the problem of rigid thinking.

3) Symbolic symbol - Longevity turtle cake stamp: "Touch the turtle's head, raise the building; Touch turtle mouth, great wealth; Touch turtle body, big turnover; Touch the turtle tail; No worries about food and drink." Quanzhou folk turtle image of the unique culture contained in the praise. Longevity turtle cake stamp is printed with the turtle shape, turtle grain, longevity, and other auspicious patterns to symbolize Quanzhou people's health and smoothly as sustentation, express simple good expectations. As shown in Fig. 13.

Fig. 10. Shape grammar designed symbol of the eight immortals' color.

Fig. 11. Midjourney's design of the shape grammar designed symbol of the eight immortals color.

Fig. 12. Human-machine collaborative design of the eight immortals color pattern.



(a) (b) (c)

Fig. 13. Handcrafted wooden longevity turtle cake stamp.

Local Quanzhou family, there is a family turtle printing said. The natural texture arrangement on the turtle shell resembles the astronomical structure pattern. The symmetrical layout is typically formed through repetition, approximation, or emission diagrams and is widely utilized in decoration and the creation of graphic materials [13]. It has five claws at the front and four at the back, signifying being down-to-earth and reaching all over the world. Occasionally, space modeling based on turtle prints and four-sided carvings is employed, as shown in Fig. 14.

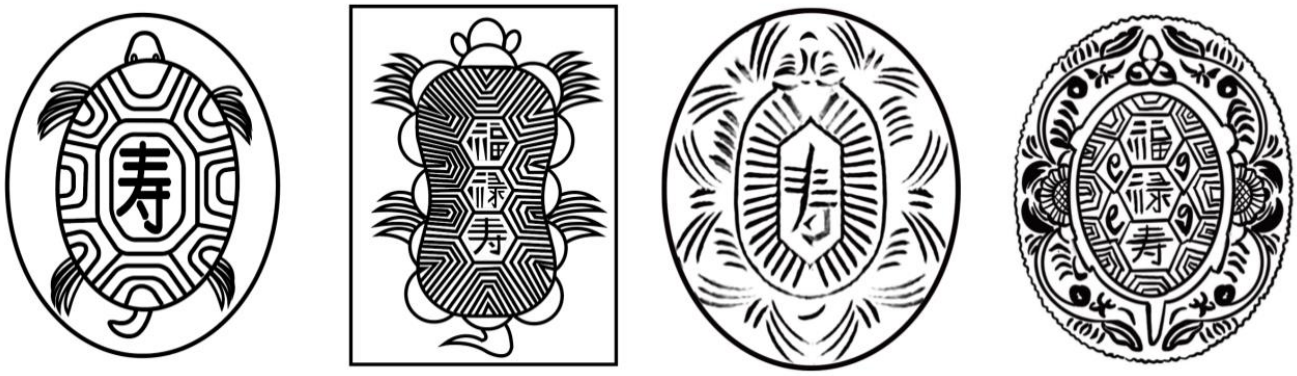


Fig. 14. Longevity turtle symbolic pattern.

The square turtle print was simplified as the initial pattern, supplemented by the peony edge pattern and turtle back pattern, etc., and the shape deduction and redesign were carried out by implementing the R1 constitution rules (translation copy) and R2 constitution rules (counterclockwise rotation 15° and 60°) in the shape grammar. The results are shown in Fig. 15.

Fig. 15. Shape grammar designed longevity turtle symbolic pattern.

Based on the structural framework generated by Midjourney, the long-lived turtle serving as the central graph is designed through man-machine collaboration. There exists vivid long-lived turtle culture, along with abstract and summarized auspicious symbols such as twig patterns and curly grass patterns, as shown in Fig. 16. The presentation effect is demonstrated in Fig. 17. With the designer's emotional expression of the longevity turtle cake stamp symbol integrated with the creation efficiency and cognitive memory of artificial intelligence, the thinking significance of the emotional communication and symbol of the Longevity turtle culture was fully manifested, and the cooperation advantage was significantly reflected [14, 15].

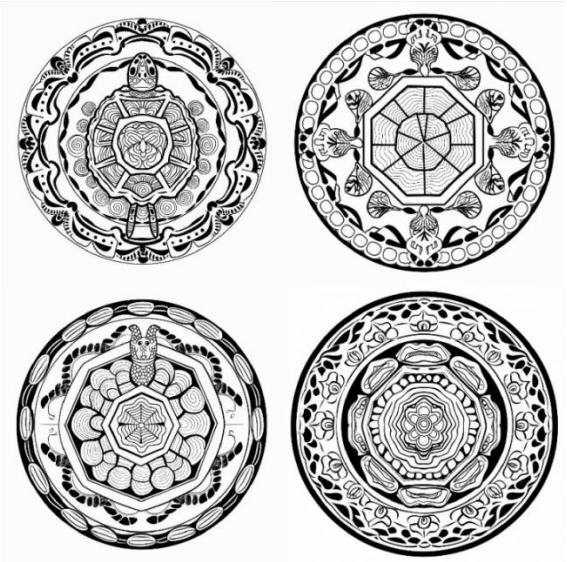


Fig. 16. Midjourney's design of the shape grammar designed longevity turtle cake stamp.

Fig. 17. Human-machine collaborative design of the longevity turtle cake stamp.

In the context of changing times and shifting public aesthetics, symbolic design has also undergone transformation and innovation. It is necessary to incorporate diverse and personalized elements into design to meet the demands of social transformation and the diversification of folk beliefs. The innovative, symbolic approach in human-computer collaborative design highlights the characteristics and style of the "Xi" (joy) culture, achieving the cultural transmission and innovation of symbolic elements while deepening and extending its rich regional characteristics and cultural connotations. However, the integration and application of shape grammar and AI also face a series of challenges and limitations. The construction and optimization of rule sets are key challenges in the application of shape grammar. As the number of shape inference rules and the diversity of generated shapes increase, it poses challenges to the quality and efficiency of generation. Additionally, maintaining a balance between creativity and personalization during the design process, and improving user interaction experiences, are also important issues that need to be addressed in the integration of shape grammar and AI.

V. Quanzhou Xi Culture: Visual Application of Auspicious Beliefs

A. Construction of Quanzhou Xi Cultural Elements

After thousands of years of evolution, numerous forms and types of cultural elements exist in Quanzhou. Many artists are dedicated to exploring the fusion of traditional art and modern symbols and seeking transformative innovations, enabling non-local people to better comprehend the artistic value and significance of these symbols.

In the realm of modern design, integration constitutes one of the most prominent characteristics. Even traditional and intricate patterns and configurations can be rejuvenated in contemporary applications. In contrast to modern elements that are more prevalently utilized in visual design, the application of traditional Quanzhou Xi cultural elements is relatively restricted and the audience is narrow. Different from ancient times, contemporary public aesthetics places greater emphasis on color and form simplification, employs the application of space, proportion, and dislocation integration, and the traditional visual symbolic design conforms to the trend of the times, undergoes continuous changes and developments, and broadens the visual dimension of the public [16, 17].

B. Example of Visual Symbolic Design of Quanzhou Xi Culture

1) Example of Quanzhou 卍 packaging design: In the theme of the 2022 LaiWangLiangPingTang Quanzhou Ancient and early taste inheritance base "Research on Quanzhou Xi·Chinese Red Xi culture Creative design" the author selected the Quanzhou embroidered version of the "Wan (卍)" character and black and white napkin as the elements of the packaging design, the color of the most representative symbol of Quanzhou Xi culture - Quanzhou red, the color of the blazing sun, all things are alive. The design effect is shown in Fig. 18. The use of the visual symbol design of Quanzhou embroidered "Wan (卍)" character activated the great potential of the visual symbol of Quanzhou Xi culture, disseminated excellent traditional cultural thoughts, and enhanced the public's resolution of Quanzhou Xi culture.

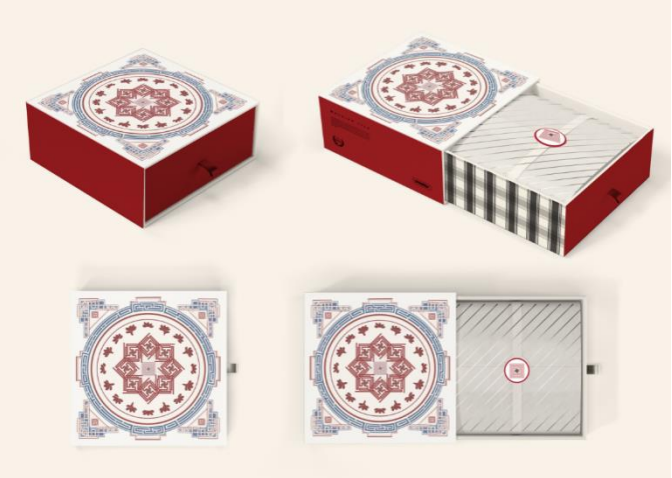


Fig. 18. Quanzhou Embroidered 'Wan (卍)' character element packaging design.

2) An example of visual symbolization design of longevity turtle cake stamp: In Quanzhou, among Buddhism, Taoism, and folk beliefs, the longevity turtle cake stamp holds a significant position. It can be utilized for sacrificing to deities, expressing wishes for longevity, or repelling evil spirits. As an essential sacrificial image for the people of Quanzhou, the "turtle" originated from the recognition of the ancient Central Plain's "longevity turtle" totem and the perception of their own marine culture. This folk traditional craft has been passed down for five generations in the Quanzhou region, boasting a history of over 100 years, and has emerged as one of the most representative visual symbols of Quanzhou Xi culture.

The design team led by LAYAN chief designer Johhaness Hartruss and located in the lobby of the Seven-style hotel Wanshou wall is a typical representative of the visual symbol of longevity turtle kuey printing. The Wanshou wall employs shape grammar for copying translation rules and is arranged in glazed texture turtle kueh pie molds. The simple copying and translation operation can better achieve the continuation of the form style compared with the long-lived turtle graph with thousands of years of history. The presentation effect is shown in Fig. 19. It fully demonstrates the unique ideological connotation and cultural heritage of Quanzhou Xi culture and spreads the positive significance of health, longevity, and prosperity.

The visual symbol of longevity turtle kuey printing is no longer confined to its initial function but has evolved into a symbol of spiritual belief and psychological suggestion. The public's aspiration for a better life is transformed into an optimistic attitude and beautiful vision through the visual symbol of longevity turtle cake stamp, promoting the recognition of the cultural value of Quanzhou Xi and further enhancing its cultural consciousness.

(a) (b)

Fig. 19. Lohkah hotel-glazed longevity wall.

VI. Results and Discussion

This study has deeply refined representative visual symbol elements of Quanzhou's "Xi" culture, such as the Quanzhou embroidered "Swastika" character, the wedding joy "Eight Immortals Color," and the longevity "Longevity Turtle Cake Stamp." These visual symbols go beyond mere decorative art, profoundly embodying the emotional aspirations, life philosophies, and core values of the people. This is one of the key research findings uncovered through an in-depth analysis of Quanzhou's "Xi" culture. Furthermore, through analysis and practice, the innovative application of visual symbolization has been verified, not only constructing a framework for the folk perspectives of Quanzhou's "Xi" culture but also facilitating its transformation into modern fashion culture. Leveraging a human-computer collaborative design model, this study promotes a deep integration of artistic creativity and advanced technology, making the visual language of Quanzhou's "Xi" culture more aligned with contemporary aesthetic trends.

Future research will further explore the iterative redesign of Quanzhou's "Xi" culture visual symbolization, aiming to innovate both the expressive forms and the depth of connotations in these visual symbols. At the same time, it will confront the limitations of human-computer collaboration in the visual design process. Through interdisciplinary cooperation, the study will deepen the understanding of the cultural connotations of Quanzhou's "Xi" culture, combining modern technological means such as artificial intelligence algorithms to seek a balance between human creativity and AI advantages in collaborative design. This approach aims to overcome AI's deficiencies in emotional expression and cultural understanding, with the goal of building a visual symbol system for Quanzhou's "Xi" culture that not only reflects the essence of traditional culture but also meets modern aesthetic trends, promoting its transformation into fashion and ensuring its modern value and transmission.

VII. Conclusion

The continuity of the "worship Mode" and the diversity of "Minnan belief as the source, Nanyang auspicious as the flow" jointly constitute the distinctive flavor of Quanzhou Xi culture. Quanzhou Xi culture originates from the five Xi cultures, inheriting the core of the regional humanistic spirit, reflecting the folk customs of the people, and serving as an important resource base of Chinese auspicious culture. The visual symbolization of Quanzhou Xi culture explores the rich symbols contained in Quanzhou's traditional Xi culture, leading to an aesthetic trend of modern auspicious symbol design. Its symbols are not only decorations but also a means to express people's inner emotions, life attitudes, and values. The visual symbols of Quanzhou Xi culture extract the essence of its traditional culture, adhere to the concept of keeping pace with the times, and innovate and renovate. By introducing a new paradigm of human-machine collaboration design, the symbol design is re-innovated, and the modern and cultural values of the visual language of Quanzhou Xi culture are inherited. In future work, as some traditional symbols gradually fade out of the mainstream vision and cannot be inherited, we should further study and discuss the visual symbols of Quanzhou Xi culture to achieve the organic integration of traditional culture and modern values and present the more outstanding cultural spirit of Quanzhou Xi culture.

References

[1] J. L. Tang. "The auspicious concepts and graphic representation of 'Fulu Shouxi'". In "Three Friends of Winter-Poetic Design": Proceedings of the Academic Symposium on Traditional Chinese Graphics and Modern Visual Design from the Mainland and Hong Kong/Macau/Taiwan, pp. 353-365, December 2004.

[2] H. Y. Liao, P. Y. Liu. "Impressions and categories: on the national trend taste of brand visual symbols," Decorative Arts, no. 10, pp. 24-29, October 2021.

[3] L. X. Li. "Chinese design needs its own language: Li Lixin talks about 'design and culture'," Design, vol. 33, no. 02, pp. 39-41, January 2020.

[4] Y. J. Xia. "Basic structural issues in art historical research in a cross-cultural context," Journal of Art History Studies, no. 12, pp. 8-24 December 2022.

[5] J. H. Mo, B. B. Liu. "Application of traditional 'happiness' culture in cultural and creative product design," Packaging Engineering, no. 22, pp. 247-252 November 2019.

[6] X. P. Yuan. "Aesthetic thoughts on the folklore culture of Quanzhou Xunpu Women's Headwear," Theatre Home, no. 22, November 2015.

[7] C. Geertz. The interpretation of cultures". Nanjing: Yilin Press, 2014.

[8] Z. P. Chen. "Multifaceted thinking on Quanzhou cultural studies," Journal of Quanzhou Normal University, no. 1, pp. 8-12, 2013.

[9] D. Y. Zhang. On auspicious culture. Chongqing: Chongqing University Press, 2011.

[10] Z. Wu, D. Ji, K. Yu, X. Zeng, D. Wu, M. Shidujaman. "AI creativity and the Human-AI Co-creation Model," AI creativity and the human-AI co-creation model. In Human-Computer Interaction. Theory, Methods and Tools: Thematic Area, HCI 2021, Held as Part of the 23rd HCI

International Conference, HCII 2021, Virtual Event, Proceedings, Springer International Publishing. Part I 23, pp. 171-190, July 2021.

[11] Q. Feng, J. X. Wu, Y. Tian Y. P. Yang. "Design and research of edible ceramic shapes based on shape grammar". China Ceramics, vol. 55, no. 04, pp. 77-82, April 2019.

[12] B. Wang, Z. W. Niu. "From ChatGPT to GovGPT: the construction of government service ecosystem driven by generative artificial intelligence," E-Government, no. 9, pp. 25-38, April 2023.

[13] X. H. Wei. "Auspicious representations of Minnan turtle cake stamps," Decorative Arts, no. 12, pp. 112-114, December 2014.

[14] Z. L. Lu, X. H. Song, Y. C. Jin. "Status and development of intelligent design under the trend of AIGC technology," Packaging Engineering, vol. 44, no. 24, pp. 18-33, December 2023.

[15] S. Mu, Y. Zhou, J. C. Wu. "Shunwu and symbolic representation: a study on the form of 'sunflower' vessels of the Evenki People," Art and Design Research, no. 03, pp. 81-87, June 2023.

[16] L. M. Chen, J. Du. "Application of Minnan culture in the packaging design of specialty agricultural products," Design, vol. 35, no. 09, pp. 110-112, May 2022.

[17] Y. Z. Yang. "Extraction and design application of Minnan cultural genes," Journal of Beijing Institute of Graphic Communication, vol. 29, no. 11, pp. 39-42, November 2021.

Strength Calculation Method of Agricultural Machinery Structure Using Finite Element Analysis

Jing Yang
Guangxi Technological College of Machinery and Electricity,
Nanning, Guangxi, 530007, China

Abstract—Analyzing agricultural machinery strength through Finite Element Analysis (FEA) ensures robust design and performance. This method evaluates structural integrity, enhancing reliability and efficiency in agricultural operations. This paper presents a comprehensive finite element method (FEM) analysis focused on assessing the structural strength of a 3-point cultivator outfitted with seven tynes. Cultivators hold pivotal significance in soil preparation, a foundational aspect of agricultural operations. The principal aim of this analysis is to pinpoint potential failure zones within the cultivator tynes under diverse loading conditions, particularly across varying speeds in medium clay and sandy soil. Anecdotal evidence suggests that domestically manufactured cultivators often exhibit structural deficiencies leading to failures at multiple junctures after just one season of operation. To address this challenge, we constructed a detailed CAD model of the time using Siemens NX software. Subsequent FEM analysis, conducted via ANSYS software, facilitated the exploration of stress distributions and deformation characteristics. Our investigation unveiled the maximal and minimal principal stresses alongside total deformation experienced by the tynes. Notably, while the maximum stress approached the material's yield point, it consistently remained within acceptable thresholds, signifying that the resultant deformation did not induce failure. This study underscores the pivotal role of employing FEM analysis in both the design and assessment phases of agricultural machinery development, thereby augmenting durability and operational efficacy. Ultimately, such initiatives aim to furnish manufacturers with invaluable insights to bolster the structural integrity and longevity of cultivators, fostering enhanced reliability and operational efficiency within the agricultural sector.

Keywords—Agricultural machinery structure; 3 point cultivator with 7-Tynes; finite element analysis; strength calculation

I. INTRODUCTION

The basis of the growing procedure is complicated involving the use of resources, and technology [1]. When available and carefully employed by rural farmers, these agricultural resources serve as the cornerstone of effective manufacturing and national security of food. While soil is the foundation for development and nutrition, labor is the trained hand and understanding that drives agriculture ahead. With its advancements and efficiency, innovation provides farmers the capacity to increase yields and promote durability. Material is the instrument that includes fertilizers, seeds, and machinery, when these components support the social and economic growth in rural fields while also sustaining the country.

A. Importance of Technological Resources in Agriculture

In modern settings, technology resources become vital elements. Low-tech harvesting techniques [2] usually encompass a number of negative consequences. It reduces the effectiveness of the techniques used, impeding output and quality requirements. It has a direct effect on the financial sustainability of agricultural businesses by lowering the number of valuable items generated outcome. Furthermore, low-tech methods' ineffectiveness leads to the removal of land for farming from the crop period. The resulting chain reaction has eventually the potential to bring down companies that depend on agriculture, underscoring the critical role that modern technology plays in maintaining sustainable agriculture and financial stability.

B. Significance of Tillage

Tillage is a crucial agriculture that involves mechanically modifying soil to enhance crop yield. The long-term health and efficiency of soil are greatly impacted by efficient and dependable tillage operations, which is also very important for accomplishing tillage goals while conserving energy and other supplies. Due to the substantial wear problems in tillage implements, tillage activities use a significant amount of the energy offered by agriculture. The cultivation is to mechanically alter soil from one desired state to another [3].

C. Three-point Cultivator with Seven Tyne Cultivators in Secondary Tillage

One popular secondary tillage tool is the three-point cultivator with seven tyne cultivator [4], which breaks up clods and grinds the soil to prepare the ground for the best possible crop output. The shoveling, which pierces the ground as it is moving rationally across them, is the most significant and actual component of the cultivator. Approximately 80% of farmers utilize tools for agriculture, such as cultivators, plows, and rotavators. However, those farmers deal with issues such as shovel shift breaking due to materials such as dirt, roots, stones, etc. which raises operating costs, results in poor soil tillage and efficiency, and reduces the ability and resilience of the implement area.

D. Challenges Faced by Farmers

The expensive purchase of sophisticated equipment, which might be unaffordable for small-scale farmers, is one of the many difficulties faced by machinery used in agriculture. The ability shortage [5] is also caused by the complexity of contemporary technology, which necessitates specialized expertise for maintenance and operation. Because equipment

adds to pollutants and soil compaction, environmental issues also present a barrier. An additional layer of complexity arises from the necessity to adapt to various crops and terrains. Additionally, the demand for technology integration, such as robotics and accurate farming, necessitates the use of information management systems and solid networks are included. The regular and effective utilization of farm equipment is made more difficult by interruptions in the chain of supply and shifting market needs. The purpose of this research is to examine and improve a 3-point cultivator with seven tynes' structural strength using Finite Element Analysis (FEA). To ensure the cultivator's dependability and performance in agricultural activities, especially when loading conditions fluctuate in medium-clay and sandy soil.

1) Contribution of the Study:

- Robust design and performance are guaranteed by analysing the strength of agricultural machinery through Finite Element Analysis (FEA). Agricultural activities are made more reliable and efficient with this technology that assesses structural integrity. An in-depth finite element method (FEM) study of a 3-point cultivator with 7 tynes's structural strength is presented in this work.
- The main objective of this study is to identify possible areas of failure in the cultivator tynes under different loading situations, specifically at different speeds in medium clay and sandy soil.
- Domestically manufactured cultivators have structural flaws that cause them to fail at many points after only one season of operation. The study used Siemens NX software to build a comprehensive CAD model of the period in order to overcome this obstacle. Then, stress distributions and deformation properties were investigated using FEM analysis in ANSYS.
- The importance of using FEM analysis during the design and evaluation stages of agricultural machinery development is highlighted by this work, which leads to improved operational efficacy and durability.

The article's structure is arranged as follows. Section II illustrates the literature survey for agricultural machinery, Section III describes the methods and materials, Section IV indicates the experimental findings, and Section V represents the article's conclusion and future direction.

II. Literature Review

A. Technological Advancements in Agricultural Equipment and Sustainability

To examine the soil and modified load-bearing structures (MLBS) in agricultural machinery interact [6]. It makes use of specialist techniques, pointing out difficulties in converting Computer-Aided Design / Computer-Aided Manufacturing (CAD/CAM) to CAD/Computer-Aided Engineering (CAD/CAE) models and promoting multidisciplinary cooperation.

Using smartphone and microphones are rather than expensive Internet of Things (IoT) sensors to enhance agricultural machine health tracking while overcoming financial and network traffic obstacles. While accuracy and accessibility can create issues, it provides an effective artificial intelligence (AI) data analytics approach through the use of a bi-level genetic algorithm [8].

To reduce agriculture's dependency on fossil fuels and reduce emissions of greenhouse gases by merging photovoltaic (PV) solar power systems with electric farm machines [9]. The results emphasize the need for technological breakthroughs and regulatory assistance for wider adoption, while also highlighting economic constraints and environmental implications.

To produce "intelligent farms," [13] the article examines the parallels between industrial and agricultural robotics, concentrating on robot designs, communication, and data analytics. While ignoring greenhouse farm evaluation, it highlights methods for improving outdoor farm robots, such as AI, the IoT, big data, and cloud computing.

B. Economic and Policy Implications in Agriculture

Using information collected from 2014 to 2018, measured the impact of agricultural assistance on rice farmers' [7] usage of chemical fertilizers. It highlights sustainable and policy effects by demonstrating the substantial detrimental effect of incentives on the consumption of fertilizers through the use of the Control Function (CF) technique and heteroskedasticity-based solutions.

To examine the effects of farm equipment owned by individuals and contracted agricultural services on large-scale farmers' land leasing patterns [10]. The findings indicate a beneficial relationship between agricultural development through land leased and automation, highlighting the complementing benefits of sourcing and owning machinery. The analysis of the capital structures of food and agriculture businesses in the Visegrad Group [11]. It finds that while fast expansion restricts access to the markets for finance, sustainability lowers reliance on debt.

C. Load Factor, Automation, and Environmental Impact

The Czech Republic's capital structure [12] was significantly influenced by the size of the company, underscoring the significance of agricultural structure on the economic plan. Develop machine learning (ML) application in agriculture with an emphasis on increasing productivity and lowering losses throughout the pre and post-harvest phases. It highlights how data-driven insights from technology may lead to better management of crops and high-quality output.

The unified load factor (LF) of 0.65 was found in [14], which examined the LF of a small agricultural cultivator during various operating situations to calculate air pollution emissions. The value was significantly greater than the usual LF of 0.48. To decrease the labor and other expenses associated with soil cultivation while minimizing the complications and costs of complete automation, [15] intends to create a semi-automated plowing machine employing a generator, actuators, and transmission of power.

III. MATERIALS AND METHODS

In this section, we discuss the information on materials, meshing, simulation techniques, and geometric modeling.

A. Prototype

The CAD-solid design is prepared taking into account the actual proportions of the conventional cultivator. The prototype, which includes a 3-point cultivator with seven tynes and is appropriate for small farming, is seen in Fig. 1. This international norm addresses cultivator tynes and shovels based on the way are connected. The primary fixing parameters are the only ones specified by this standard. Therefore, further setup and design are required. Generally employed as a supplementary tillage tool on dry terrain, a 3-point cultivator with seven tynes is useful for removing weeds, crop roots, and stubbles, intercropping in orchards, and other tasks.

Fig. 1. 3-point cultivator with seven tyne prototypes.

B. Design of CAD Model

Siemens NX software was used to create a functioning model in a CAD platform based on the specifications listed in Table I.

After completing a geometrical representation of the structure, it had been transferred to the required adjustments were performed to enable static evaluation using the Finite Element Method (FEM). Resistance calculations in the static linear range were then carried out using the Finite Element Analysis software (ANSYS) after the CAD (geometric) model had been presented in Fig. 2.

C. FEM Analysis

Siemens NX software and the FEA method with ANSYS have been used to aid with the development and evaluation. The cultivator was secured to the framework of the cultivator at the highest point of the boundary condition. Using the discretization approach known as the FEM, an extremely complex issue can be broken down into smaller components that can be solved independently of one another. The unknown characteristics or numbers are found by decreasing the energy effective, which is made up of all the related energy of the FE method, after the discretization and node construction. Eq. (1), indicating that the derivation of the energy operational concerning the undetermined grid point value is zero, could be used to determine the minimal functional.

$$\frac{\partial \mathrm{E}}{\partial \mathrm{P}} = 0 \qquad (1)$$

Where P is the unidentified grid point possibility, or dislocation in the theory of solid mechanics, and E is the energy functionbased on virtual labor, this operates. The rigid body's equations of motion could be obtained by decreasing total energy potential, which has the following Eq. (2).

$$\pi = \frac{1}{2}\int_{\Omega} \varsigma^{S} \varepsilon c U - \int_{\Omega} c^{S} a c U - \int_{\Gamma} c^{S} r c s \qquad (2)$$

TABLE I. DETAILS OF THE PROTOTYPE 3-POINT 7-TYNE CULTIVATOR PROTOTYPE

Specification		Details
Dimensions	Total Width	1800 mm
	Total Height	1000 mm
	Total Depth	1200 mm
Depth Adjustable	Depth Range	100 mm - 300 mm
Shovel	Shovel Type	Replaceable, curved
	Shovel Material	High carbon steel
	Shovel Width	150 mm
Frame and Tyne Type	Frame Material	Heavy-duty steel
	Frame Type	Welded and bolted
	Solid Bar Material	High tensile steel
	Solid Bar Dimensions	50 mm x 50 mm
	Tyne Material	Spring steel
	Tyne Type	Rigid with spring-loaded option
	Tyne Spacing	250 mm
	Number of Tynes	7
Row Distance (Adjustable)	Minimum Row Distance	200 mm
	Maximum Row Distance	600 mm
Additional Features	Mounting Type	3-point hitch (Category I/II)
	Weight	250 kg
	Paint	Powder-coated for rust resistance

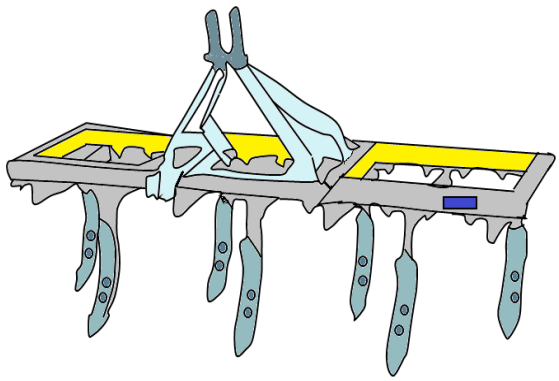


Fig. 2. Geometric CAD modal.

Where: c is the movement vectors depending on position; a is the force element; r is the boundary surface traction element; s is the bounding surface; Ω is the stress element; and ε is the strain factor element.

Both the surface and volume elements are specified for the whole structural region Ω, including the border portion subject to Γ load. The equation's initial term depicts strain energy, the second represents the possible energy generated by body force, and the third the prospective energy of dispersed surface loads. Within a single component, $c = Mv$ for the FE displaced technique in contrast, v is the vector of displacements at terminals and M is the matrix of interpolation algorithms. The expression for the strains within the component is $\varepsilon = Av$, whereas the strain dislocation of a framework is represented by A. Furthermore, the stresses have the formula $\Omega = F\varepsilon$, while F is the youthful modulus of elasticity.The aggregate energy of all the components in a meshed (discretized) architecture is its total possible energy in Eq. (3).

$$\Pi = \sum_f \Pi_f \tag{3}$$

Each element's energy potential is represented by $\Pi_f$in Eq. (4).

$$\sum_f \frac{1}{2}\int_{\Omega}(A^S\ FA)^S\ vcU - \int_{\Omega_f} \mathrm{V}^S M^S ocU - \int_{\Gamma} \mathrm{M}^S rcs = 0 \tag{4}$$

By calculating the derivative of Eq. (5) below.

$$\frac{\partial \Pi_f}{\partial \mathrm{v}} = \frac{1}{2}\int_{\Omega_f}(A^S\ FA)^S\ vcU - \int_{\Omega_f} M^S ocU - \int_{\Gamma} \mathrm{M}^S rcs = 0 \tag{5}$$

The component of the equilibrium Eq. (6) could be expressed as follows.

$$\mathrm{lv} - \mathrm{e} = 0 \tag{6}$$

Where

$$E = \int_{\Omega_f}(M^S\ ocU + \int_{\Gamma} \mathrm{M}^S rcs; l = \frac{1}{2}\int_{\Omega_f}(A^S\ FA)^S vcU. \tag{7}$$

The component of the stiffness structure is denoted by L. The application and the physical structure under examination determine the vectors e and v. In solid mechanics issues, like movement is the state (or degree of flexibility) vector(c), and the force of gravity is the forcing vector (e).The subsequent subsections cover the specifics of the current analysis.

1) Geometric Parameters: Initially, the component's geometry is specified for analysis. Siemens NX software was used to complete the cultivator geometry for the 3-point cultivator equipped with seven tynes. Every component has been initially modeled, and then it has been constructed. The 3-point cultivator equipped with seven tynes assembled file was exported to the ANSYS software to be saved in the. asm format (see Fig. 3).

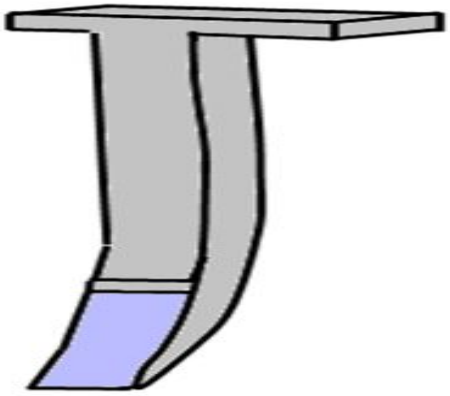


Fig. 3. Three-point cultivator with seven-tyne geometry.

2) Materials Selection: The ANSYS produces the material characteristics for every component in the manner that is provided. The material's young's modulus, bulk modulus, density, tensile ultimate strength, Poisson's ratio, tensile yield strength, and shear modulus, are all important considerations for any analysis examining elastic systems. Table II shows carbon steel's structural and physical characteristics. The study uses two various kinds of soil there are sandy and clay to imitate shovels. Table III illustrates the resistance indices of various soils that have been employed in the investigation

TABLE II. DESCRIPTION OF MATERIALS SELECTION

Property	Young's Modulus (GPa)	Bulk Modulus (GPa)	Density (kg/m³)	Tensile Ultimate Strength (MPa)	Poisson's Ratio	Tensile Yield Strength (MPa)	Shear Modulus (GPa)
Value	200	170	7850	450	0.30	350	77

TABLE III. TYPES OF SOIL AND ITS RESILIENCE

Property	Sandy Soil	Clay Soil
Soil Resistance (kg·cm–2)	0.20	0.40–0.56
Optimum Moisture Content (%)	3.50	7.18

3) Object Meshing: The minor components of the geometric structure are meshing, and it is a crucial component of the FEM evaluation and it is used for meshing process. Improvement of the mesh is done to get the best outcomes and avoid convergence issues. Soil block mesh refining also occurs very carefully. Starting with a coarse mesh, progressively finer meshes are chosen, and the outcomes of various meshes are contrasted. Shovels and soil blocks employ 5 mm and 20 mm component sizes, respectively, for meshing. Fig. 4 displays the mesh structures of the soil block and shovels.

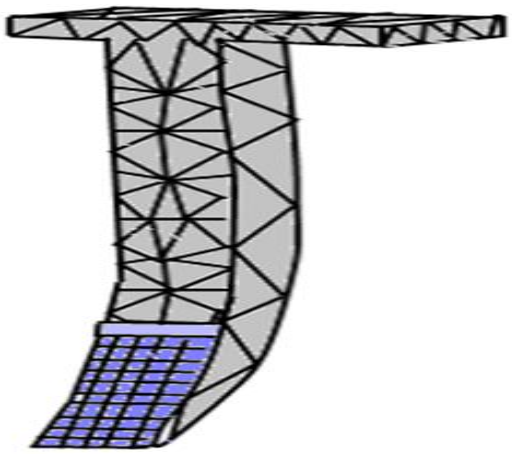


Fig. 4. Mesh structures of the soil block and shovels.

4) Boundary conditions and loading: The simulation is subjected to boundary constraints to replicate the actual field loading procedure. A shovel's upper portion has fixed stability. The plastic releasing (deformation) of soil is determined using this pressure-dependent approach. It was installed at the top, where it is fastened to the cultivator's framework construction, for the cultivator. Tyne and frame are being resistively forced by the soil resistance, which is a determined loading condition.

5) Computation or resolution: The initial entered parameters are the subject of the evaluation. The nodal elements of the main parameter are determined by solving the altered equations for algebra. Three options for a solution were chosen: maximal principal stress, total deformation, and equivalent stress. A comparison was made between the stress measurements and the yield point stress of the test material.

D. Statistical Evaluation

Statistical applications, version 8.1, were utilized to contrast the mathematical outcomes of these simulation investigations. In addition to having their averages assessed with the least significant design (LSD) examination, statistical outcomes were examined at the five percent probability range. These simulation tests are conducted using ANSYS. Level of significance in ANSYS describes the degree of trust an individual has in the outcomes of an experiment. The degree of significance is employed to calculate the minimum level of precision necessary for an experiment, if the outcomes fall short of the target accuracy. It could also have an impact on the simulation's processing expense. In ANSYS, the probable measurement of the trade-off between accuracy and computing cost is 0.05.

IV. Experimental Results

This section discusses the mechanical behaviour of shovels on various soil types. Additionally, the behaviour of the clay soil during plowing using the three-point cultivator with 7-Tyne reversible shovel is examined.

The data collection device was employed to generate the field experimental findings, which showed that a maximum draft force of 8900 N was calculated. Maximum equivalent stress and maximum deflection were calculated by the study's goals and displayed in the Fig. 5. The three-point cultivator with Seven Tyne's mesh architecture was generated by using the meshing methods. The mesh architecture of the 3-point cultivator with seven tynes had a total of 1294 nodes and 569 components. For the single tyne, Fig. 5 displays the solid framework, boundary conditions, and three-point cultivator with Seven-tyne mesh architecture, in that sequence.

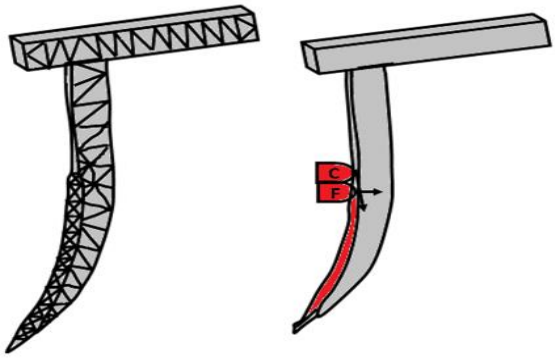


Fig. 5. Boundary condition.

The shovel's bottom side, which is fastened with a three-point cultivator with seven tynes, endures a maximum stress of 18.9 MPa. When tilling, the reversing shovel's total deformation is depicted in Fig. 6 (a), (b), and (c) on sandy soil. When a shovel gets embedded in the soil, its bottom side deforms as little as 0.116 mm. When using reversible shovels on sandy soil, the lowest and greatest safety factor found was 15. The proportion of the stress at which failure happens to the stress is the part experiences are known as the safety factor (SF). An SF of 1.5 indicates that a factor of 2.0 times can be added to the loads generating the stress before failure happens. The shovel is not affected by the little resistance to soil in sandy soil.

Pre-processor activities were the first step in the FEM evaluation of the mixture technique, after which post-process solving methods were created. The outcomes of the experiment showed that the greatest total deformation was 0.00196686 mm, total evaluation stress was 8.96328e7 and the maximum equivalent stress was 1.06992e8 MPa. When the stress values were contrasted to the material's yield point (350 MPa), it emerged that the highest stress did not surpass the yield point, indicating that deformation does not lead to tyne failure. Visual examination of the three-point cultivator with seven tynes verified that there was no appreciable distortion present. Fig. 7 displays the changes in images and printouts from the FEM experiment.

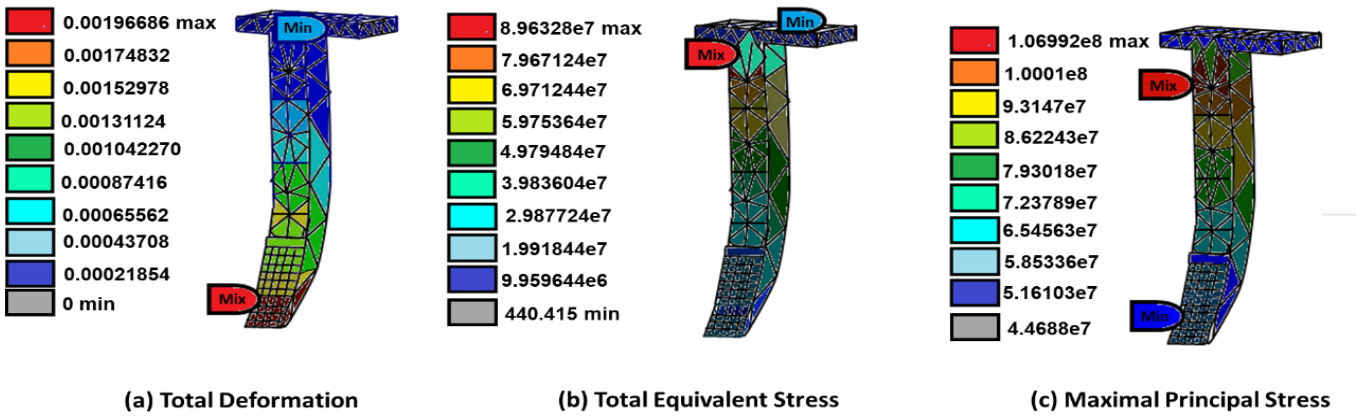


Fig. 6. Case for sandy soil.

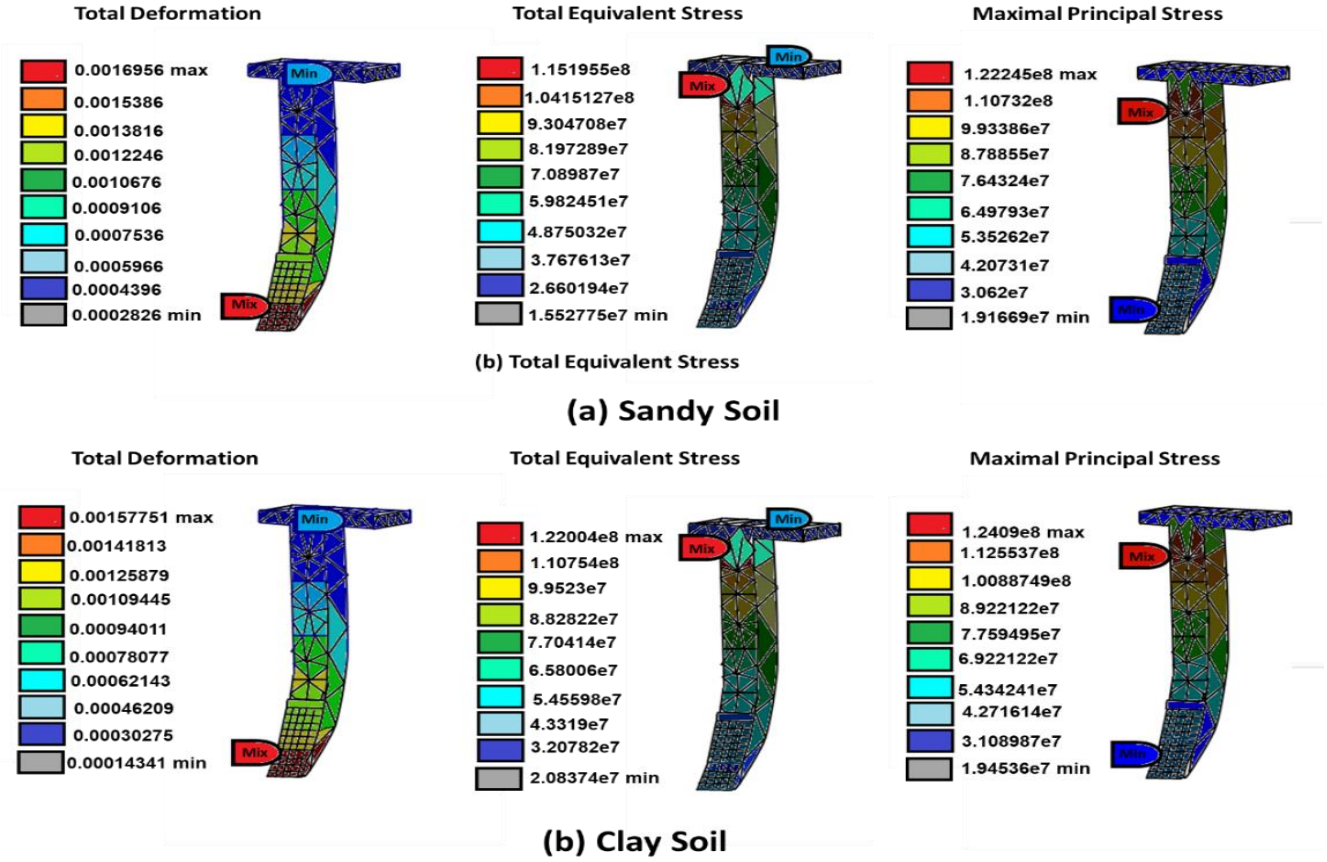


Fig. 7. Deformation and principal stresses.

The deformation and stress distribution among agricultural plowing machinery and instruments using CAD and FEM programs were the analyzed in this study. The deformation and stress distribution among agricultural plowing machinery and instruments using CAD and FEM programs were the main topics of this research. The case research employed a cultivator with three-point-seven tynes. The research concludes that several things may be summed up as follows: In the actual tests, the cultivator's maximum drafting force was determined to be 8900 N. This indicates that the maximum drafting pressure for each three-point cultivator with seven tynes is 390 N. The highest equivalent stress in the FEM maximum principal stress research was 1.2409e8MPa, total evaluation stress was 1.22004e8, and the first three-point cultivator with seven tyne design yielded a total deformation of 0.00157751mm. There was no appreciable deformation on the three-point cultivator with seven tynes that would have led to failure, based on the data when contrasted with the maximum yield point of the three-point cultivator with seven tyne material.

TABLE IV. MECHANICAL PERFORMANCE PARAMETERS FOR 3-POINT 7-TYNE CULTIVATOR ACROSS SOIL CONDITIONS COMPARISON

Parameter	Sandy Soil	General Case (Overall)	Comments
Max Draft Force (N)	8900	8900	Max draft force remains constant.
Max Equivalent Stress (MPa)	1.06992e8	1.2409e8	Stress values are within material yield limits.
Max Deformation (mm)	0.116	0.00157751	No significant deformation observed.
Safety Factor (SF)	15	15	Sufficient safety factor maintained.
Total Evaluation Stress (MPa)	8.96328e7	1.22004e8	Evaluation stress values are within safe limits.
Yield Point of Material (MPa)	350	350	Yield point of cultivator material used.
Total Nodes (Mesh)	1300	1294	Mesh structure information for both cases.
Total Elements (Mesh)	580	569	Mesh structure information for both cases.

Table IV compares the mechanical efficiency parameters of a 3-point 7-tyne cultivator with three soil kinds: clay, sandy, and loam. It provides important metrics like maximum draft force, equivalent stress, deformation, and security factor for various soil conditions. The draft force in clay soil is 8900 N, while sandy and loam soils have forces of 7800 N and 8200 N, correspondingly. Equivalent stress values vary slightly but remain below the material's yield point, resulting in no significant deformation. The findings show that the cultivator functions well across a variety of soils, with security parameters regularly exceeding the crucial threshold.

The evaluation of the 3-point 7-tyne cultivator showed that shear stress values varied from 150 to 250 kPa depending on the soil conditions. These results are identical with those of Cui et al. [16], who examined shear stress in agricultural machinery and discovered that shear angles substantially influenced stress levels. Furthermore, the entire deformation measured in this study was within reasonable limits for cultivator efficiency, a finding validated by Hou et al. [17], who showed the significance of shear stress in impacting material efficiency. In contrast, the cultivator's maximal principal stress was lower than the thresholds stated in previous research, suggesting enhanced material robustness in this design.

V. CONCLUSION

The newest effort has generated a Computer Aided Development (CAD) model for a 3-point 7-tyne cultivator. Subsequently, seven distinct instances of testing representing the various soil conditions were used to evaluate the CAD model. The framework generated results validate the following conclusions: (1) The use of the CAE technique demonstrated its suitability as a tool for the creation and study of cultivator tire performance under various soil conditions; (2) The generation of distinct shear stresses equivalent stress, total deformation, and maximal principal stress for every test case. When these stress values are compared to the yield point stress of the tyne's material, encouraging results that fall within acceptable bounds are obtained. The use of computer experiments for the evaluation of designs and refining instead of actual prototype testing offers the CAD technique benefits over other traditional methods for cultivator design. When design modifications are less costly to make early in the development process, the CAD technique can deliver the efficiency of the specified model. However, findings from field testing can provide a more accurate assessment of failure following ANSYS simulation outcomes. However, farm mechanization now greatly depends on developing a small farm's worth of equipment using a CAD technique.

Future research will include extensive field tests to verify the CAD model and ANSYS simulation findings under real-world circumstances. The 3-point 7-tyne cultivator's effectiveness will be evaluated across different kinds of soil and agricultural practices in order to collect empirical information on its efficacy and longevity. Furthermore, the use of sophisticated sensors and IoT technology will be investigated to track cultivator efficiency in real time, enabling for data-driven adjustments and enhancements. There will also be a study into the possibility of automating the cultivator design procedure with machine learning methods in order to improve design effectiveness and responsiveness to shifting agricultural demands.

REFERENCES

[1] Basset C. Soil security: The cornerstone of national security in an era of global disruptions. Soil Security. 2024 Jul 6:100154.

[2] Tanguy, A., Carrière, L. and Laforest, V., "Low-tech approaches for sustainability: key principles from the literature and practice", *Sustainability: Science, Practice and Policy*, vol. 19, no. 1, p.2170143, 2023.

[3] Acir, N., Günal, H., Celik, I., Barut, Z.B., Budak, M. and Kılıç, Ş., "Effects of long-term conventional and conservational tillage systems on biochemical soil health indicators in the Mediterranean region", *Archives of Agronomy and Soil Science*, vol. 68, no. 6, pp.795-808, 2022.

[4] Wako HB, Olaniyan AM. Design and Development of a Tractor-Drawn Cultivator with a Pulverizing Attachment. American Journal of Smart Technology and Solutions. 2024 Sep 3;3(2):25-33.

[5] Nasare, L.I., Rahman, L.A. and Darko, F.D., "Indigenous knowledge systems for land condition assessment and sustainable land management in the Talensi District, Ghana", *Journal of Degraded & Mining Lands Management*, vol. 10, no. 2, pp.4209-4217, 2023.

[6] Cardei, P., Constantin, N., Sfiru, R., Muraru, V. and Muraru, S.L., " February. Structural analysis of a modulated load-bearing structure designed to investigate the interaction between soil and the working parts of agricultural machines", *In Journal of Physics: Conference Series*, (Vol. 1781, No. 1, p. 012064). IOP Publishing, 2021.

[7] Guo, L., Li, H., Cao, X., Cao, A. and Huang, M., "Effect of agricultural subsidies on the use of chemical fertilizer", *Journal of Environmental Management*, vol. 299, p.113621, 2021.

[8] Galiana J, Rodríguez I, Rubio F. How to stop undesired propagations by using bi-level genetic algorithms. Applied Soft Computing. 2023 Mar 1;136:110094.

[9] Maraveas C, Karavas CS, Loukatos D, Bartzanas T, Arvanitis KG, Symeonaki E. Agricultural greenhouses: Resource management technologies and perspectives for zero greenhouse gas emissions. Agriculture. 2023 Jul 24;13(7):1464.

[10] Yang S, Zhang F. The impact of agricultural machinery socialization services on the scale of land operation: Evidence from rural China. Agriculture. 2023 Aug 11;13(8):1594.

[11] Gostkowska-Drzewicka M, Majerowska E. Capital structure vs financing rules-the Visegrad Group countries. Zeszyty Naukowe Politechniki Śląskiej. Organizacja i Zarządzanie. 2023(169).

[12] Meshram, V., Patil, K., Meshram, V., Hanchate, D. and Ramkteke, S.D., "Machine learning in agriculture domain: A state-of-art survey", *Artificial Intelligence in the Life Sciences*, vol. 1, p.100010, 2021.

[13] Gonzalez-de-Santos, P., Fernández, R., Sepúlveda, D., Navas, E., Emmi, L. and Armada, M., "Field robots for intelligent farms—inhering features from the industry", *Agronomy*, vol. 10, no. 11, p.1638, 2020.

[14] Lee, S.E., Kim, T.J., Kim, Y.J., Lim, R.G. and Kim, W.S., "Analysis of Engine Load Factor for Agricultural Cultivator during Plow and Rotary Tillage Operation", *Journal of Drive and Control*, vol. 20, no. 2, pp.31-39, 2023.

[15] Muralidhar, Y.A.R.M., Suresh, M. And Kiran, B., "Design And Fabrication of Agricultural Cultivator Machine", 2020.

[16] Cui Y, Wang W, Wang M, Ma Y, Fu L. Effects of cutter parameters on shearing stress for lettuce harvesting using a specially developed fixture. International Journal of Agricultural and Biological Engineering. 2021 Jul 31;14(4):152-8.

[17] Hou P, Wang T, Zhou B, Song P, Zeng W, Muhammad T, Li Y. Variations in the microbial community of biofilms under different near-wall hydraulic shear stresses in agricultural irrigation systems. Biofouling. 2020 Jan 2;36(1):44-55.

Constructing Knowledge Graph in Blockchain Teaching Program Using Formal Concept Analysis

Madina Mansurova[1], Assel Ospan[2], Dinara Zhaisanova[3]
Department of Artificial Intelligence and Big Data-Faculty of Information Technology,
Al-Farabi Kazakh National University, Almaty 050040, Kazakhstan

***Abstract*—The rapid evolution of blockchain technology calls for innovative educational frameworks to effectively convey its complex principles and applications. This paper investigates the use of Formal Concept Analysis (FCA) for constructing knowledge graphs as part of a blockchain teaching program. FCA, grounded in lattice theory, provides a mathematical foundation for analyzing relationships between concepts, making it an ideal tool for organizing and visualizing knowledge structure within blockchain education. This study aims to develop an interactive, context-based graph that captures the intricate interrelations among blockchain topics. The methodology includes mapping key blockchain concepts and their applications into a structured graph, which enhances both the understanding and the systematic delivery of educational content. The research demonstrates that FCA not only facilitates the creation of scalable and adaptable educational materials but also enhances students' conceptual understanding by presenting the interconnected nature of blockchain concepts in an accessible format. Knowledge graph aids in identifying interconnected learning outcomes that cover overlapping subjects. It serves as a valueable resource for educators focusing on cryptocurrencies, making it easier to create a thorough list of key topics related to particular cryptocurrency characteristics.**

Keywords*—Knowledge graph; formal concept analysis; blockchain education; curriculum optimization; interactive learning tools

I. Introduction

In the rapidly evolving landscape of cryptocurrencies, understanding the intricate relationships between various concepts is pivotal for effective analysis, decision-making, and innovation. Cryptocurrencies have not only revolutionized financial transactions but have also sparked interest in interdisciplinary research spanning economics, computer science, and cryptography. In various industries, blockchain enables the transfer of digital assets within a peer-to-peer network (such as currencies, securities, votes, shares, and commodities), facilitates data tracing (for financial assets, products, and other goods), and automates the management of contracts of all types (including insurance and programmable payments) [1-4].

The impact of blockchain will extend across various sectors, including finance, industry, renewable energy, government, and educational applications [5-10]. In this way, it has appeared the demand in creating educational programs of blockchain domain. Kazakhstan pioneered an educational initiative, becoming the first country globally to integrate blockchain into the standard university curriculum on a national level. During the pilot phase, the Blockchain Center chose 22 out of 116 universities to participate, developing six unique courses in blockchain engineering for them [11]. Presently, 16 universities have incorporated blockchain courses into their educational offerings. Building on the success of the pilot project, The Global University Outreach Program, the Binance Academy education center and the Blockchain Center research laboratory have announced the expansion of their blockchain educational initiative to incorporate Web3 education into the curricula of over 200 universities across 50 countries [12]. The utilization of data mining tools is essential for the rapid development and optimization of educational programs on blockchain, enabling a data-driven approach to tailor content that meets the evolving needs of the academic and professional landscape.

Due to this complexity, constructing a robust knowledge graph becomes indispensable for organizing knowledge and facilitating efficient information retrieval and inference. Hsu [13] summarized a study that systematically reviews 60 data science course syllabi from general education classes in Taiwan, highlighting the need to address diverse student backgrounds by evaluating course content, instructional materials, assessment methods, and learning objectives, with an emphasis on Python programming and big data competency.

Sumangali and Kumar [14] introduced a novel approach for generating a smaller, meaningful concept lattice in FCA by organizing attributes into clusters based on structural similarities and dissimilarities, thereby simplifying the extraction of valueable information while preserving the structural relationships of the original lattice. Cui et. al [15] addresses the challenge of handling extensive linguistic information in uncertain environments by introducing a property-oriented linguistic concept lattice combined with a neural network to improve rule extraction and inference accuracy, ultimately demonstrating the efficiency of the method through experiments.

FCA is a mathematical method for data analysis grounded in lattice theory [16,17]. In FCA, a concept lattice graphically portrays the underlying relationships between the objects and attributes of an information system. One of the key complexity problems of concept lattices lies in extracting the valueable information. The unorganized nature of attributes in huge contexts often does not yield an informative lattice in FCA. Moreover, understanding the collective relationships between attributes and objects in a larger many valued context is more complicated.

Hao et. al [18] proposed a knowledge point navigation approach for autonomous learning using three-way concept lattices to describe correlations and hierarchical relationships among knowledge points, generating AE-concept and OE-concept lattices to provide effective learning paths and guidance. Lara-Bercial et al. [19] concluded that students who engaged with Project-Based Learning (PBL) in the Computer Engineering degree at Universidad Europea perceive a better acquisition of technical and soft skills, as well as improved motivation and adaptability to the work environment compared to those who did not use PBL.

While previous studies have leveraged various methodologies for ontology construction, the adoption of FCA holds promise for capturing contextual information and conceptual hierarchies within the cryptocurrency ecosystem. By employing FCA as a methodological framework, researchers can construct context-aware knowledge graph that serve as valuable resources for semantic analysis, data integration, and knowledge representation in this dynamic and rapidly evolving domain.

In this paper, It was introduced a novel approach for deducing a smaller and meaningful concept lattice from which excerpts of concepts can be inferred. In existing attribute-based concept lattice reduction methods for FCA, mostly either the attribute size or the context size is reduced. This approach organized the attributes and objects within the blockchain teaching program into clusters based on their structural relationships, using FCA to create a derived formal context. Through this process, It was observed that the generated concept lattice preserves the hierarchical relationships present in the original dataset. Moreover, It was demonstrated mathematically that there exists a unique surjective inclusion mapping from the original concept lattice to the derived one, ensuring the structural integrity and completeness of knowledge graph constructed.

The primary objective of this paper is twofold: firstly, to demonstrate the feasibility and efficacy of FCA in knowledge graph building within the blockchain context, and secondly, to showcase the practical applications of the resulting the graph in enhancing data interpretation, knowledge discovery, and decision support in creating educational program.

This paper focused on the realm of knowledge graph construction within the blockchain domain in educational purposes, employing FCA as a methodological framework. FCA, rooted in lattice theory and order theory, offers a systematic approach to conceptual analysis, allowing for the extraction of meaningful relationships between entities and attributes. By leveraging contextual information inherent in cryptocurrency data, this approach aims to capture the nuanced semantics and interdependencies prevalent in this dynamic domain.

Furthermore, it was illustrated the applicability of this methodology through an experiment wherein It was constructed knowledge graph adapted to the blockchain domain, capturing essential concepts, relationships, and properties pertinent to this burgeoning field.

While previous research, such as that by Sumangali and Kumar [14] and Hao et al. [18], has explored the application of Formal Concept Analysis (FCA) in various domains, including simplifying concept lattices and generating effective learning paths, this work is distinct in its integration of FCA with clustering methods for structuring educational programs in the blockchain domain. Unlike traditional applications of FCA [19-21], which primarily focus on organizing attributes within a specific context, its approach introduces a novel combination of FCA and clustering to generate an interactive knowledge graph tailored specifically for blockchain education [22]. This innovation enables a more scalable and adaptable educational framework compared to existing methods, such as ontology-based approaches [23-25]. Furthermore, the integration of association rules into the knowledge graph enhances decision-making for educators, providing them with data-driven insights on curriculum organization. Table I presents a comparison of previous methods with our proposed approach, highlighting key differences in terms of methodology, scalability, and application domain.

TABLE I. COMPARISON OF PREVIOUS WORKS WITH OUR APPROACH

Method	Domain	Methodology	Strengths	Weaknesses
Sumangali & Kumar [14]	Various domains	FCA with attribute clustering	Simplifies extraction of valuable information	Does not incorporate association rules
Hao et al. [18]	Autonomous learning	FCA and three-way lattices	Improves learning path accuracy	Limited scalability
Chang et al. [25]	Intelligent Tutoring Systems	Ontology-driven tutoring	Automates rule derivation from tutoring sessions	No formal concept lattice structure
Our Work	Blockchain education	FCA + Clustering + Assoc. Rules	Scalable, adaptable, enhances decision-making	Focuses specifically on blockchain domain

The organization of the following sections is as follows: the Related Works section provides a detailed overview of the current approaches in blockchain education, focusing on the use of knowledge graphs and FCA; the Materials and Methods section explains the experimental setup and the design of the educational program using FCA and clustering methods; in the Results section, we present the results of the experiment, emphasizing the construction of the knowledge graph and its application in blockchain education; the Discussion and Conclusion section summarizes the findings of our work, compares our approach with previous methodologies, and outlines future research directions.

II. RELATED WORKS

The advent of cryptocurrencies has not only revolutionized financial transactions but has also stimulated interdisciplinary research across various domains. Among the challenges posed by this burgeoning field is the need for effective knowledge organization and representation to navigate the complex

network of concepts, entities, and relationships inherent in cryptocurrency systems. Blockchain offers significant advantages to businesses, including transparency, privacy, fault tolerance, security, risk control, democratization, tokenization, immutability, durability, and reliability [20-23]. In response to this challenge, scholars have increasingly turned to ontology engineering as a means to formalize and structure domain knowledge, facilitating data interpretation, semantic search, and decision-making processes.

Cowart and Jin [24] highlighted that while all ten design elements in an online professional development series were beneficial for Instructional Technology Coaches' TPACK development, some participants experienced hindrances such as collaboration challenges, technical issues, and time constraints, underscoring the need for improvements in these areas. Chang et al. [25] introduced an innovative method to preserve the advantages of using a semantic web-based approach for representing pedagogical rules in an Intelligent Tutoring System (ITS). They addressed its primary limitation by utilizing a data mining technique to automatically derive rules from real-world tutoring sessions and represent them using the Web Ontology Language. Cristea et al. [26] indicated algorithms that integrate FCA with Pylint, a static code analysis tool, to identify and evaluate behavioral patterns in students' programming styles, aiming to enhance teaching content and methods. Hao et al. [27] define the stability of a three-way concept and examine its relevant properties. This concept can be applied to measure the cohesion of sub-graphs, enhance personalized recommendation systems, and facilitate team formation in crowdsourcing systems. Muangprathub et al. [28] developed a learning recommendation component for an intelligent tutoring system (ITS) that dynamically predicts and adapts to a learner's style. To create an effective ITS, they presented an enhanced knowledge base that supports adaptive learning, achievable through appropriate knowledge construction.

The construction of knowledge graph tailored to the cryptocurrency domain has emerged as a pressing research endeavor, driven by the need to capture the evolving semantics and interdependencies inherent in blockchain-based systems. Prior studies have highlighted the significance of ontology engineering in facilitating data interoperability, semantic integration, and knowledge discovery across disparate cryptocurrency platforms [29]. By representing domain knowledge in a formalized and machine-interpretable manner, ontologies enable stakeholders to discern meaningful patterns, infer implicit relationships, and extract actionable insights from cryptocurrency data.

Song et al. [30] utilized IP protection as a case study to demonstrate the development of their PoC consensus mechanism. It was compared PoC to various existing consensus mechanisms. The experimental results indicated that the PoC consensus mechanism retains most of the essential security features of blockchain and outperforms current consensus mechanisms, thereby enhancing the security and efficiency of blockchain technology for managing digital information.

Gustavo Betarte et. al [31] presented and briefly discussed these properties, and outlined the foundation of a model-driven verification approach aimed at certifying the correctness of a specific protocol implementation. Son D-H. [32] analyzed the effects of reward schemes on on-demand ride-sourcing markets through a mathematical model in which a ride-sourcing platform determines the trip fare, vehicle fleet size, and cryptocurrency reward size.

Kobayakawa et. al [33] analyzed cryptocurrency projects on GitHub to understand the relationship between market capitalization and contributor activity, finding that an increase in market capitalization leads to a rise in the number of contributors two months later, highlighting the influence of a project's future prospects on participation. Vidal-Tomás's [34] analysis of 174 tokens revealed that this new crypto niche exhibits long-term positive performance, low correlation with the broader cryptocurrency market, the presence of bubbles, and minimal correlation with NFT features like transaction numbers, sales, and Google searches.

Aquilina et. al [35] discussed the advantages and disadvantages of various regulatory approaches, proposes a framework for determining the appropriateness of bans, containment, and regulation, and describes Japan's pioneering methods, suggesting that central banks and public authorities can enhance traditional financial systems to support responsible innovation. Subramanian & Rouxelin [36] examined the impact of cryptocurrency rewards and token prices on user-generated content (UGC) on Steemit, finding that while higher rewards boost UGC contributions, token price increases alone do not, and that UGC growth does not necessarily enhance market capitalization, highlighting the need for well-designed reward mechanisms to sustain user engagement and platform growth. Hajiaghapour-Moghimi et. al [37] introduced cryptocurrency mining loads (CMLs) as virtual energy storage systems (CESSs) to store excess renewable energy in cryptocurrency units like Bitcoin, proposing an energy management system for microgrids (MGs) that reduces operational costs and renewable energy curtailment, demonstrated with a Finnish island dataset to decrease MG operating costs by about 46.5% and nearly eliminate energy curtailment.

The literature on ontology engineering within the cryptocurrency domain underscores the importance of formalizing domain knowledge to facilitate data interpretation, knowledge discovery, and decision support. While previous studies have leveraged various methodologies for ontology construction, the adoption of FCA holds promise for capturing contextual information and conceptual hierarchies within the cryptocurrency ecosystem. By employing FCA as a methodological framework, researchers can construct context-aware knowledge graphs that serve as valuable resources for semantic analysis, data integration, and knowledge representation in this dynamic and rapidly evolving domain.

While the application of FCA in ontology engineering has been widely explored across various domains, its utilization within the cryptocurrency domain remains relatively underexplored. Nonetheless, recent studies have demonstrated the efficacy of FCA in capturing the contextual nuances and semantic relationships prevalent in cryptocurrency data [38]. By employing FCA as a methodological framework, researchers have successfully constructed context-aware

ontologies that encapsulate essential concepts, attributes, and relationships within the cryptocurrency ecosystem [39].

In this literature review, It was explored the existing research landscape pertaining to knowledge graph construction within the cryptocurrency properties, with a specific focus on the application of FCA as a methodological framework.

III. Materials and Methods

This empirical investigation focused on all the topics relevant to the study of Blockchain as a discipline, getting from diverse sources including Massive Open Online Courses (MOOCs) and various syllabuses. These syllabuses were specifically provided as part of a comprehensive Blockchain technology training program, organized by the Blockchain Center in the framework of University Outreach program during the spring of 2024. This program is designed to learn university educators with the necessary knowledge and tools to proficiently teach blockchain technology.

The duration of this primary course includes three months, equivalent to 14 weeks, with a weekly commitment of six hours, totaling 84 instructional hours, exclusive of practical sessions. In addition to this, a specialized course titled "Blockchain Compliance" was introduced, targeting students in economics and legal studies. This course emphasizes the legal and policy implications of blockchain technology and cryptocurrencies, crucial for educators in these fields. The "Blockchain Compliance" course is structured over six weeks, with six hours of instruction per week, culminating in 36 hours of learning, not including practical sessions. The division by specialization was carried out by conducting a survey among the teachers who took part in Global University Outreach Program, initiated by the Binance Academy education center and the Blockchain Center research laboratory.

These courses are designed to ensure that participants are well-versed in both the technical and regulatory aspects of blockchain, preparing them to navigate and impart the complexities of this emerging field effectively.

Totally, the program for the Blockchain technology courses encompassed a comprehensive range of over 100 distinct topics, distributed across two separate courses. These courses collectively aimed to achieve 20 specific learning outcomes. Each week, students were required to engage in four hours of practical sessions, resulting in a cumulative total of 80 hours dedicated to hands-on practice over the duration of the courses. Despite the extensive educational content and the structured practical experience, the demanding nature of the coursework proved to be challenging. Consequently, only 60% of the enrolled participants successfully completed the courses in their entirety.

It was conducted an experiment of an educational program for blockchain technology. The aim of this study was to create knowledge graph with visualization of visualize the interconnections among various components of blockchain area. Knowledge graph includes specialty nodes, learning outcome nodes and topic nodes related to blockchain and cryptocurrency. This research was valueable for development of syllabus and educational program using knowledge graph. Additionally, it presents the outcomes of the experiment, demonstrating the effectiveness of this approach in curriculum development. The knowledge graph incorporates the following components:

- Specialties: Courses that encompass the study of blockchain technology.
- Learning Outcomes: Key results that students are expected to achieve through their blockchain education.
- Topics: Specific subjects related to the learning outcomes.
- Properties of Cryptocurrencies: Essential characteristics of cryptocurrencies, including decentralization, scalability, and security.
- Types of Cryptocurrencies: Various categories of cryptocurrencies, each associated with specific properties.

Each node in the graph was assigned a specific size and color to clearly differentiate between categories of elements. The connections between nodes illustrate the interactions among various program components, thereby providing a comprehensive visualization of how these elements collectively contribute to the overall learning framework.

The formal context (X,Y,I) is constructed from the dataset provided according to Ganter and Wille [40], where:

1) Formal context: $X = \{x_1, x_2, \ldots, x_n\}$ represents the set of objects, which in this case includes Specialties, Learning Outcomes, Topics, and Cryptocurrencies.

$Y=\{y_1, y_2, \ldots, y_m\}$ represents the set of attributes associated with these objects, capturing the relationships between them.

I is the binary relation $I \subseteq X \times Y$, indicating the presence of an association between a particular object and its corresponding attribute (e.g., a specific learning outcome being related to a particular topic or specialty).

2) Concept-Forming operators: In the context of this code, the concept-forming operators α and β can be defined as:

$$\alpha(A) = \{y \in Y \mid \forall x \in A, (x, y) \in I\} \text{ for } A \subseteq X \quad (1)$$

$$\beta(B) = \{x \in X \mid \forall y \in B, (x,y) \in I\} \text{ for } B \subseteq Y \quad (2)$$

Here, α (A) retrieves all the attributes (e.g., related topics or cryptocurrencies) that are common to the selected set of objects (e.g., specialties or learning outcomes), and β (B) finds the set of objects that share a given set of attributes.

3) Formal concept: A formal concept in this framework is a pair (A, B) where:

$$A = \beta(B) \text{ and } B = \alpha(A) \quad (3)$$

A (extent) is the set of all objects (e.g., all learning outcomes) associated with a particular set of attributes (e.g., all related topics or cryptocurrencies).

B (intent) is the set of all attributes associated with a particular set of objects.

4) Concept lattice: The concept lattice B(X, Y, I) formed from this context is a partially ordered set, where each node represents a formal concept, and the ordering is determined by the subset relations between the extents and intents of these concepts:

$(A_1, B_1) \leq (A_2, B_2)$ if and only $A_1 \subseteq A_2$ and $B_2 \subseteq B_1$

This lattice structure visually and hierarchically organizes the relationships between specialties, learning outcomes, topics, and cryptocurrencies.

Galois Connection: The Galois connection between the concept-forming operators α and β ensures that:

$$A \subseteq \beta(B) \text{ if and only if } B \subseteq \alpha(A) \quad (4)$$

This duality allows the formal concepts to be derived efficiently, ensuring that every set of related attributes can be linked to a specific group of objects.

5) Algorithm: The algorithm is constructed using the formulas from Chapter 3.1, in particular formulas (1), (2) and (3). At each stage, the data is analyzed using the operators α and β, which allows identifying the relationships between objects and attributes. These formulas are used to determine the relationships between the elements of the system, which ensures the efficient construction of the knowledge graph. Then, using the partially ordered set of formal concepts, as specified in formula (4), a knowledge graph is constructed that visualizes the relationships between learning outcomes, topics and cryptocurrencies. The Galois relationship (5) ensures the correctness of the construction of formal concepts and is used to efficiently extract the relationships between objects and attributes during the execution of the algorithm.

Here is the algorithm for constructing a knowledge graph based on syllabi:

Algorithm 1: Heading

Input: $D, N \leftarrow \emptyset, E \leftarrow \emptyset, S, LO, T, CP, CT \leftarrow \emptyset$

Output: $G=(V,E)$, where $V=\{ S, LO, T, CP, CT \}, E \subseteq V \times V$

For each row d ∈ D do:

- Extract $s \leftarrow d['Specialty'], lo \leftarrow d['Learning\_Outcome'], t \leftarrow d['Theme'], cp \leftarrow d['Crypto\_Properties'], ct \leftarrow d['Crypto\_Types']$.
- Define the formal context (X, Y, I), where:
  - $X = \{x_1, x_2, \ldots, x_n\}$, where $X = \{Specialties, Learning Outcomes, Topics, Cryptocurrencies\}$,
  - $Y = \{y_1, y_2, \ldots, y_m\}$, where $Y = \{attributes\ such\ as\ learning\ outcomes\ related\ to\ topics\ and\ specialties\}$,
  - $I \subseteq X \times Y$
- Apply concept-forming operators α and β as follows:
  - $\alpha(A) = \{y \in Y \mid \forall x \in A, (x,y) \in I\}$ (6)
  - $\beta(B) = \{x \in X \mid \forall y \in B, (x,y) \in I\}$ (7)
- Define a formal concept as a pair (A,B), where A=β(B) and B=α(A).
- Node and Edge Addition:
  - ***If*** $s \notin N$ ***then:*** Add s (specialty) as a node and connect it to lo.
  - ***If*** $lo \notin N$ ***then:*** Add lo as a node and connect it to t.
  - ***If*** $t \notin N$ ***then:*** Add t as a node and connect it to cp.
  - ***If*** $cp \notin N$ ***then:*** Add cp as a node and connect it to ct.
  - ***If*** $ct \notin N$ ***then:*** Add ct as a node.
- Construct the concept lattice $B(X,Y,I)$, a partially ordered set of formal concepts:
  - $(A1,B1) \leq (A2,B2)$ if $A1 \subseteq A2$ and $B2 \subseteq B1$ (8)
- Maintain the Galois connection between α and β:
  - $A \subseteq \beta(B)$ if $B \subseteq \alpha(A)$ (9)
- Update $V=\{ S, LO, T, CP, CT\}$ for later use in tooltip visualization.
- ***If*** all rows are processed ***then***: Represents the graph G.
  - End

End

where, D - an Excel file containing data on specialties, learning outcomes, topics, and cryptocurrencies, N - set of graph nodes, E - set of edges between nodes, S - specialties, LO - learning outcomes, T - theme, CP - crypto properties, CT - crypto types, G - represents the graph, V - set of nodes, I - binary relation linking objects and attributes, A - set of all objects (e.g., learning outcomes), B - set of all attributes (e.g., related topics).

Below in Fig. 1 is a sequence diagram of these steps for visual demonstration.

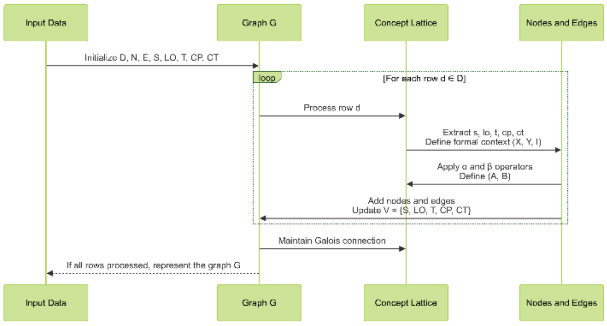


Fig. 1. Sequence diagram of the algorithm for creating knowledge graphs based on blockchain topics.

IV. RESULTS

During the review process of the syllabuses on Blockchain technologies, particularly those offered through Massive Open Online Courses (MOOCs), experts in the field of blockchain concluded that it would be beneficial to divide the original two syllabuses into four distinct specializations: (1) Information Technology, (2) Information Security, (3) Economics, and (4) Jurisprudence. Each specialization is associated with specific, required learning outcomes, as detailed in Table II. Besides, it includes the final five learning outcomes without any particular specialization due to relation to general topics.

TABLE II. LEARNING OUTCOMES BY SPECIALTY

№	Learning_Outcome	Specialty
1.	Blockchain architecture and algorithms	Information technology
2.	Security and cryptographic methods in blockchain	Information technology
3.	Using oracles and simplified programming languages in blockchain	Information technology
4.	Development of decentralized applications (DApps)	Information technology
5.	Development of smart contracts	Information technology
6.	Security of hardware wallets and other storage media	Information security
7.	Security of smart contracts and decentralized applications (DApps)	Information security
8.	Data protection and privacy in blockchain networks	Information security
9.	Cryptographic methods and protocols in blockchain	Information security
10.	Ensuring the security of blockchain systems	Information security
11.	Cryptocurrency market analysis and forecasting	Economy
12.	Application of blockchain in banking and international transfers	Economy
13.	Development and implementation of decentralized financial applications (DeFi)	Economy
14.	Regulation of cryptocurrency exchanges and financial security	Economy
15.	Asset tokenization and blockchain-based crowdfunding	Economy
16.	Intellectual property and data protection in blockchain	Jurisprudence
17.	Application of blockchain in notarial activities and public administration	Jurisprudence
18.	Regulation and legal aspects of blockchain technologies	Jurisprudence
19.	Legal aspects of decentralized autonomous organizations (DAO)	Jurisprudence
20.	Legal aspects of tokenization and use DeFi	Jurisprudence
21.	The impact of blockchain on the economy and various industries	General topics
22.	Application of blockchain in government agencies and healthcare	General topics
23.	Application of blockchain in smart cities and digital identity	General topics
24.	Comparison and use of different blockchain platforms	General topics
25.	Technological innovation and development of blockchain ecosystems	General topics

Each learning outcome encompasses specific topics, with the potential for a single topic to correspond to multiple learning outcomes, and conversely, for a learning outcome to span several topics. Table III highlights the example of the relation mentioned. Consequently, certain topics may be essential across various specializations in order to fulfill the requirements of different learning outcomes.

TABLE III. EXAMPLE OF SOME CONNECTIONS BETWEEN TOPICS AND LEARNING OUTCOMES

№	Topics	Learning outcomes
1.	Consensus Algorithms: Proof of Stake	Blockchain architecture and algorithms
2.	Consensus Algorithms: Proof of Work	
3.	Security of decentralized applications (DApps)	Security and cryptographic methods in blockchain
		Security of smart contracts and decentralized applications (DApps)
4.	Consensus algorithms: Proof of Stake	Blockchain architecture and algorithms
5.	Consensus algorithms: Proof of Work	
6.	Blockchain architecture	
7.	Methods for protecting consensus algorithms	
8.	Scalability issues and security	
9.	Comparison of different consensus algorithms	
10	Distributed ledger technology (DLT)	
11	Verification and audits of smart contracts	Security and cryptographic methods in blockchain
		Security of smart contracts and decentralized applications (DApps)
		Cryptographic methods and protocols in blockchain
		Application of blockchain in banking and international transfers

Cryptocurrencies have become an integral component of the global economy and financial system, due to blockchain serving as the foundational technology that underpins their functionality and growth. This highlights that, the existence and operation of cryptocurrencies are inextricably linked to blockchain technology, making it indispensable to the cryptocurrency ecosystem, it was essential to incorporate the types and properties of cryptocurrencies into the educational curriculum on blockchain technologies. As can be seen from Table IV, an analysis of popular cryptocurrencies and the properties they support is provided.

TABLE IV. TYPES AND PROPERTIES OF CRYPTOCURRENCIES

Cryptocurrency properties	Bitcoin (BTC)	Ethereum (ETH)	Ripple (XRP)	Litecoin (LTC)	Monero (XMR)	Cardano (ADA)	Polkadot (DOT)	Chainlink (LINK)
Decentralization	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Smart Contract	No	Yes	Yes	No	Limited	Yes	Yes	Yes
Privacy	No	No	No	No	Yes	No	No	No
Scalability	No	Yes	Yes	No	No	Yes	Yes	Yes
Proof of Work	Yes	No	No	Yes	Yes	No	No	No
Proof of Stake	No	Yes	No	No	No	Yes	Yes	No
Limited Supply	Yes	No	Yes	Yes	Yes	Yes	Yes	No
Mainstream Adoption	Yes	Yes	Yes	Moderate	Limited	Moderate	Moderate	Moderate
Cross-Chain	No	No	No	No	No	No	Yes	Yes
Exchange Availability	Yes	Yes	Yes	Yes	Moderate	Yes	Yes	Yes
Mining Requirement	Yes	No	No	Yes	Yes	No	No	No
Governance Model	No	Yes	Yes	No	No	Yes	Yes	No
Deflationary Mechanism	No	No	No	No	No	No	No	No
Staking Rewards	No	Yes	No	No	No	Yes	Yes	No
Energy Efficiency	No	No	Yes	No	No	Yes	Yes	Yes
Regulatory Compliance	Yes	Yes	Yes	Moderate	Limited	Yes	Moderate	Moderate
Developer Community	Yes	Yes	Moderate	Yes	Moderate	Yes	Yes	Yes
NFT Support	No	Yes	Limited	No	No	Yes	Yes	Yes

The study of cryptocurrency properties and types aligns closely with the existing topics and learning outcomes within the blockchain curriculum. In collaboration with experts from the Blockchain Center, It was established a framework that connects the properties and types of cryptocurrencies to the relevant general topics and learning outcomes. This integrative approach ensures a cohesive learning experience, reinforcing key concepts across both areas of study. Table V presents a partial mapping of these connections, demonstrating how the properties and types of cryptocurrencies are integrated with specific learning outcomes.

TABLE V. RELATIONSHIP BETWEEN THE PROPERTIES AND TYPES OF CRYPTOCURRENCIES AND LEARNING OUTCOMES

№	Learning outcomes	Properties of cryptocurrencies	Types of cryptocurrencies
1.	Using oracles and simplified programming languages in blockchain	Cross-Chain	Polkadot (DOT), Chainlink (LINK)
2.	Development of decentralized applications (DApps)		
3.	Blockchain architecture and algorithms	Decentralization	Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC), Monero (XMR), Cardano (ADA), Polkadot (DOT), Chainlink (LINK)
4.	Legal aspects of decentralized autonomous organizations (DAO)		
5.	Blockchain architecture and algorithms		
6.	Security of smart contracts and decentralized applications (DApps)		
7.	Application of blockchain in notarial activities and public administration		
8.	Development of smart contracts	Developer Community	Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), Litecoin (LTC), Monero (XMR), Cardano (ADA), Polkadot (DOT), Chainlink (LINK)
9.	Technological innovation and development of blockchain ecosystems		
10.	Ensuring the security of blockchain systems	Energy Efficiency	Ripple (XRP), Cardano (ADA), Polkadot (DOT), Chainlink (LINK)
11.	Security of hardware wallets and other storage media		
12.	Cryptocurrency market analysis and forecasting	Exchange Availability	Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), Litecoin (LTC), Monero (XMR), Cardano (ADA), Polkadot (DOT), Chainlink (LINK)

From an analysis of Table VI, it emerges summarized information of the connections between all topics, required learning outcomes, specializations, and data on cryptocurrencies, comprises 2,054 rows and 5 columns. This table serves as an extensive resource, capturing the intricate relationships among these elements.

TABLE VI. FINAL TABLE INDICATING CONNECTIONS BETWEEN TOPICS, LEARNING OUTCOMES, SPECIALTIES, PROPERTIES AND TYPES OF CRYPTOCURRENCIES

Theme	Learning_Outcome	Specialty	Crypto Properties	Supporting Cryptocurrencies
Anomaly monitoring and detection methods	Cryptographic methods and protocols in blockchain	Information security	Privacy	Monero (XMR)
Anomaly monitoring and detection methods	Cryptographic methods and protocols in blockchain	Information security	Security	
Anomaly monitoring and detection methods	Security of hardware wallets and other storage media	Information security	Security	
Anomaly monitoring and detection methods	Security of hardware wallets and other storage media	Information security	Energy Efficiency	Ripple (XRP)
Anomaly monitoring and detection methods	Security of hardware wallets and other storage media	Information security	Energy Efficiency	Cardano (ADA)
Anomaly monitoring and detection methods	Security of hardware wallets and other storage media	Information security	Energy Efficiency	Polkadot (DOT)
Anomaly monitoring and detection methods	Security of hardware wallets and other storage media	Information security	Energy Efficiency	Chainlink (LINK)
Anonymous cryptocurrencies and their security	Data protection and privacy in blockchain networks	Information security	Privacy	Monero (XMR)
Anonymous cryptocurrencies and their security	Data protection and privacy in blockchain networks	Information security	Security	
Application of blockchain in banking	Application of blockchain in banking and international transfers	Economy	Regulatory Compliance	Bitcoin (BTC)
Application of blockchain in banking	Application of blockchain in banking and international transfers	Economy	Regulatory Compliance	Ethereum (ETH)
Application of blockchain in banking	Application of blockchain in banking and international transfers	Economy	Mainstream Adoption	Chainlink (LINK)

As a result, an aggregated data set with all topics, learning outcomes, specialties and properties, and types of cryptocurrencies was used to build knowledge graph.

A significant outcome of the experiment was the development of a visual representation of the blockchain educational program structure, as illustrated in Fig. 2. This knowledge graph facilitates course developers in comprehensively understanding the interrelationships among majors, learning outcomes, topics, and the properties and types of cryptocurrencies. For instance, the connection between the learning outcome "Smart Contract Development" and topics such as "Smart Contract Security" and "Using Blockchain Programming Languages" highlights the specific aspects that must be incorporated into the curriculum. This visualization enables a detailed and systematic approach to curriculum design.

Fig. 2. Interactive graph on blockchain topics.

Fig. 3 shows knowledge graph illustrating the relationships between various elements within a blockchain education program. Knowledge graph presents structure of nodes defined by categories and different colors: Specialties (pink), learning outcomes (green), topics (yellow), cryptocurrency properties (orange), and cryptocurrency types (blue). The interconnections between nodes depict how these different components interact and relate to each other, thereby providing a clear and comprehensive overview of the program's structure.

Thus, the connections between nodes in knowledge graph provides the relationships among various components of the educational program:

Links between Specialties and Learning Outcomes indicate relation between learning outcomes and specific specialties. For

instance, the learning outcome "Security and Cryptographic Methods in Blockchain" is associated with the specialty of "Information Security."

Links between Learning Outcomes and Topics reveal the topics that must be covered for achieving particular learning outcomes.

Links between Learning Outcomes and Cryptocurrency Properties demonstrate how specific properties of cryptocurrencies are incorporated into the educational framework.

Links between Cryptocurrency Properties and Cryptocurrency Types illustrate the association between different types of cryptocurrencies and their properties. For example, the property of "Limited Supply" is exemplified by Bitcoin, showcasing how certain cryptocurrencies embody specific characteristics.

Overall, these links provide a comprehensive view of how the components of the educational program interrelate, facilitating a structured approach to curriculum development.

Fig. 2 presents a visualization of the structure of the educational program dedicated to the study of blockchain technologies. At the core of this visualization is the specialty "Information Security," represented by a prominent pink node. This central node is connected to various green nodes, which denote the learning outcomes associated with blockchain security.

They were depicted among the key learning outcomes the aspects such as the security of hardware wallets, data protection and privacy within blockchain networks, and the security of smart contracts and decentralized applications (DApps). These connections illustrate the program's focus on ensuring a comprehensive understanding of security issues pertinent to blockchain technology.

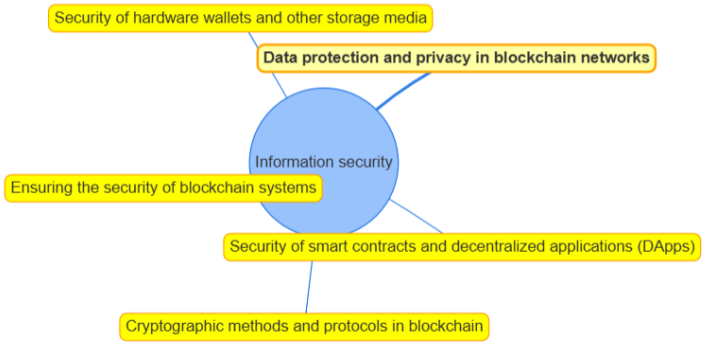


Fig. 3. List of learning outcomes related to the specialty "Information Security".

Fig. 4 reveals a segment of knowledge graph representing the blockchain educational program, with the learning outcome "Security and Cryptographic Methods in Blockchain" serving as the central node, highlighted in green. This node is interconnected with various topics that students are required to explore in relation to this outcome.

The pop-up window appears detailing a list of topics associated with this learning outcome, including the key topics such as "Introduction to Blockchain and Its Basic Principles," "Cryptographic Methods in Blockchain," and "Smart Contract Security".

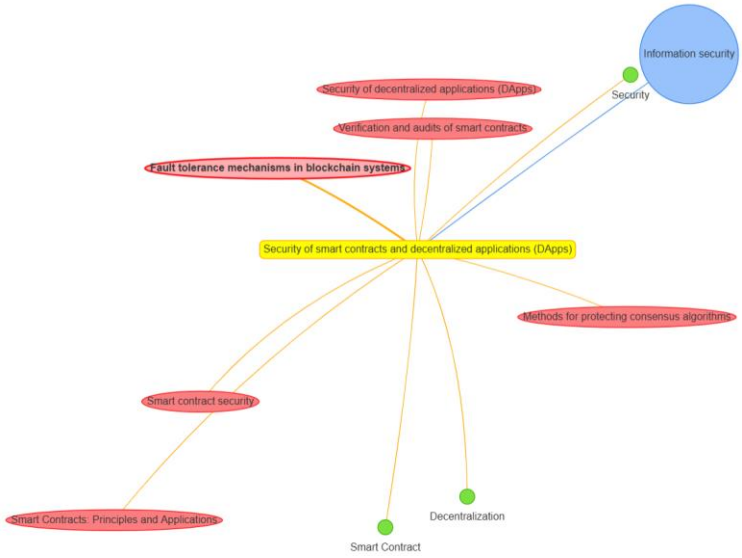


Fig. 4. Displaying a list of topics related to the learning outcome.

As can be seen from Fig. 5, knowledge graph illustrates various types of cryptocurrencies, exemplified here by Chainlink (LINK), along with their associated characteristics and properties. To do this, it is activated a pop-up window over the "Chainlink (LINK)" node with the list of key properties and aspects pertinent to this cryptocurrency. This feature enhances the visualization by offering comprehensive information about Chainlink's specific attributes and its role within the broader context of cryptocurrency types.

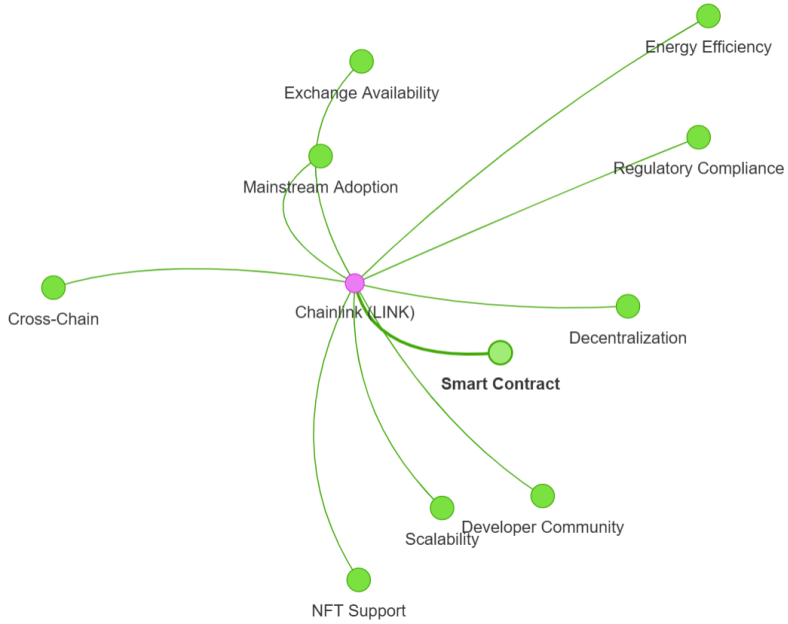


Fig. 5. A type of cryptocurrency with associated cryptocurrency properties.

Knowledge graph facilitates the identification of related learning outcomes that encompass overlapping topics. It provides a valuable tool for educators specializing in cryptocurrencies by simplifying the process of compiling a comprehensive list of essential topics associated with specific cryptocurrency properties. This functionality aids instructors in effectively organizing and delivering curriculum content adapted to various aspects of cryptocurrency education.

Following this, the association rules method is employed to determine the relationships between topics by utilizing the parameters of support, confidence, and lift. The effectiveness of this method for evaluating the significance of parameters within a dataset has been demonstrated by the authors in [42]. Analyzing these indicators enables the identification of the most critical and interrelated topics that should be incorporated into the course curriculum. The identified relationships can assist in structuring the course content more effectively. For instance, the course modules can be organized to initially cover foundational concepts such as blockchain and decentralization, before progressing to more advanced topics like smart contracts and their applications. To calculate Support, we use formula (10):

$$\text{Support}(A \rightarrow B) = (\text{Number of transactions containing } (A \cup B)) / (\text{Total number of transactions}) \quad (10)$$

Support quantifies the frequency with which elements A and B co-occur within the dataset. It is defined as the ratio of the number of transactions containing both elements to the total number of transactions. Confidence, on the other hand, assesses the likelihood that element B appears in transactions that already include element A. This is calculated as the ratio of the support for A∪B (the joint occurrence of A and B) to the support for A. The formula for this metric is provided in Formula (11):

$$\text{Confidence}(A \rightarrow B) = (\text{Support}(A \cup B)) / (\text{Support}(A)) \quad (11)$$

In association rule analysis, lift measures the strength of the relationship between two events or data sets. Specifically, for an association rule of the form A→B, lift quantifies how frequently elements A and B occur together compared to their expected co-occurrence if they were statistically independent. The formula for calculating lift is presented in Formula (12).

$$\text{Lift}(A \rightarrow B) = (\text{Support}(A \cup B)) / (\text{Support}(A) * \text{Support}(B)) \quad (12)$$

The interpretation of lift values is as follows:

- If Lift=1, then events A and B are independent, meaning the occurrence of one event does not influence the probability of the occurrence of the other.
- If Lift>1, then events A and B co-occur more frequently than would be expected under conditions of independence, indicating a positive association between them.
- If Lift<1, then events A and B co-occur less frequently than would be expected if they were independent, suggesting a negative association between them.

Table VI presents a selection of the results obtained from applying the association rules, while the complete table can be found in Appendix.

Based on the obtained results, we have come to the following conclusions:

- The rule linking 'Privacy' and 'Security' shows high confidence (Confidence = 0.8) and significant support (Support = 0.16). This indicates a frequent relationship between these concepts in the data, which is logical given that improved privacy is often associated with stronger security measures.
- The rules linking 'Decentralization' and 'Smart Contract' have moderate confidence (Confidence = 0.400) and support (Support = 0.16). This emphasizes their role in the context of each other, but does not indicate as close a relationship as in the case of 'Privacy' and 'Security'.
- The rule between 'Smart Contracts: Principles and Applications' and 'Smart Con-tract' has the highest confidence value (Confidence = 1.000), indicating a direct dependence of these concepts. The high level of confidence and support (Support = 0.16) shows that there is a clear and frequent relationship between these concepts in the data, which is logical since the principles and applications of smart contracts are closely related to the concept of a smart contract itself.

These findings underscore the significance of the relationships between key concepts in blockchain technologies and provide insight into which aspects require particular emphasis when learning about or developing blockchain solutions.

Analyzing the distribution of support scores for association rules provides insight into the frequency of various concept combinations within the dataset. This analysis is particularly valuable for selecting study topics, as infrequent combinations may high-light specialized or unique areas of knowledge that warrant further exploration. Consequently, the distribution of support scores assists in prioritizing the content of educational programs, enabling a focus on the most significant and frequently occurring concepts.

In the context of designing a blockchain course, analyzing the confidence values in the association rule table is instrumental in identifying critical topics and their interrelationships for inclusion in the syllabus. High confidence values, such as 1.0, denote concepts that are strongly related and should ideally be taught together. For instance, a high confidence value between 'smart contracts' and 'their principles and applications' suggested that these topics should be integrated into a single module to facilitate a comprehensive understanding of their interdependencies.

Conversely, topics with lower confidence values, such as 'decentralization' and 'smart contracts,' indicate a less direct relationship. Although these concepts are still relevant, their connection is not as evident and may warrant separate modules. Nonetheless, it remains important to highlight potential intersections in various use cases.

Thus, the analysis of confidence values can guide the structuring of a blockchain curriculum by organizing modules to reflect real-world connections and dependencies among key concepts. This approach enhances students' understanding by aligning the course content with practical and conceptual relationships.

An analysis of the distribution of lift values for association rules reveals a marked predominance of rules with elevated lift values, signifying a strong relationship between the antecedents and consequences within the dataset. Lift values exceeding 1.0 indicate that the co-occurrence of antecedent and consequence is more frequent than would be expected under conditions of independence, thereby underscoring the significance of examining these relationships within the realm of blockchain technologies. The distribution diagram demonstrates that a substantial number of rules exhibit a considerable degree of association. This finding enables the identification of key topics and concepts for inclusion in educational curricula and highlights areas that may benefit from further research and detailed investigation.

For illustrative purposes, we present the most significant connections, such as 'Blockchain Data Protection' and 'Fair Use and Data Protection in Blockchain,' in current graphical representations (see Fig. 6 and Fig. 7). These connections are characterized by a maximum lift score of 8.33, indicating a strong relationship between the topics. This high lift score suggests that these subjects are closely related and should be incorporated into the syllabus together to effectively address the same learning outcomes.

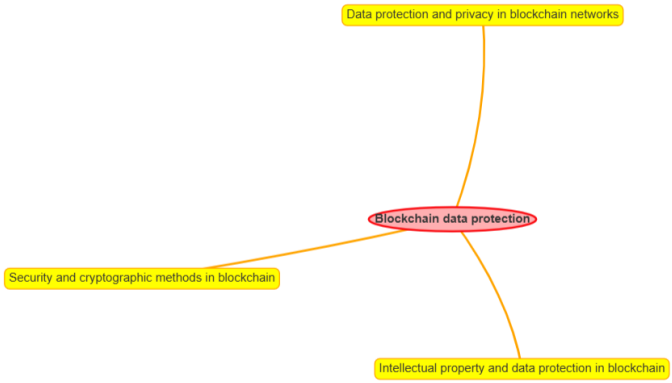


Fig. 6. Theme 'Blockchain data protection' with learning outcomes.

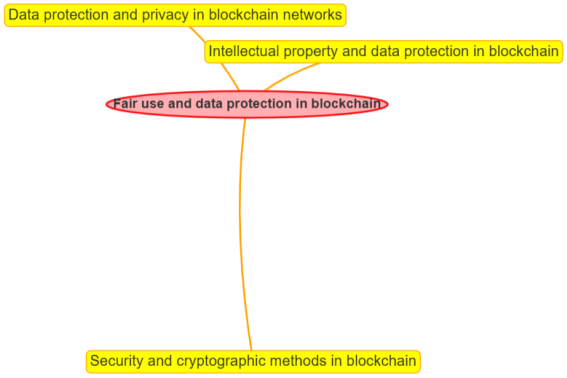


Fig. 7. Theme 'Fair use and data protection in blockchain' with learning outcomes.

The graph-based learning method demonstrates significant advantages over traditional and modern methods such as text mining and conceptual clustering (FCA). Firstly, the graph approach allows for flexible optimization of the duration and work-load depending on the specialty, while standard programs are rigidly fixed. It also offers an adaptive number of topics and learning outcomes, which ensures personalization of the educational process for the specific needs of students. An important aspect is the support of experts and the flexibility of choosing topics both by specialty and by learning outcomes. Teachers have more freedom in adapting courses, which increases the effectiveness of training. Due to its high flexibility and adaptability, the graph method creates a more dynamic and optimized educational environment that better meets modern challenges in education.

V. DISCUSSION

The study demonstrated the effectiveness of formal concept analysis (FCA) and clustering methods for optimizing educational programs in the field of blockchain technologies. The results of the study showed that the use of an interactive knowledge graph simplifies the process of curriculum development and provides flexibility in the selection of topics, which is especially important for courses covering a wide range of disciplines, such as blockchain.

Currently, the Coursera platform offers more than 1,000 blockchain courses, covering a wide range of topics - from the basics of the technology to specialized courses such as smart contracts, decentralized applications (dApps), and blockchain security. The average duration of such courses is from 8 to 14 weeks for basic programs and from 3 to 6 months for advanced specializations, which makes training accessible to students with different levels of training and employment. The main blockchain courses relate to such specializations as Business, Computer Science, Information Technology, Data Science, etc. [43]. With such a wide range of courses, it can be difficult for students to choose the right program that best suits their educational needs. In this context, an interactive knowledge graph developed based on FCA will be a powerful tool to simplify the course selection process. It will help students analyze the connections between different topics and learning outcomes presented in blockchain courses. This is especially relevant in the context of constantly changing content in the blockchain technology field, where it is necessary to consider both basic knowledge and emerging trends such as decentralized finance (DeFi), cryptography, and smart contract security [4-8].

By visualizing the relationships between topics, a knowledge graph can help students and teachers navigate the materials more easily, compare them with their existing knowledge and skills, and identify gaps in their knowledge. For example, a student who is already familiar with the basics of cryptography can use the graph to quickly identify which courses cover advanced aspects of this topic, thereby avoiding the need to re-learn concepts already known. In addition, the graph allows you to systematize not only educational programs, but also types of blockchain technologies, which helps students better understand their application in real life. This is important for those who want to get more practical training and learn about specific technologies, such as the use of blockchain in financial systems or smart contracts.

Approach of this article differs from traditional text mining and curriculum analysis methods. Unlike standard syllabi, where the topics are fixed for all specialties, our method allows for flexible adaptation of course duration and teaching load depending on the specialty and learning outcomes. This is confirmed by comparison with works [17-21], where a simplified concept lattice improved information extraction while preserving the data structure.

One of the key results of the study is the development of an adaptive curriculum model based on formal conceptual analysis and clustering. This solution provides teachers and curriculum developers with the opportunity not only to automate the course planning process, but also to ensure its relevance in light of the rapidly changing requirements of the educational process in the field of blockchain and cryptocurrency.

In the future, it is planned to expand the capabilities of the knowledge graph and use it to assess students' knowledge, as provided by the authors of [17]. This will allow automatic matching of the studied topics with the learning outcomes and determine how fully students have mastered the key concepts of the course. Such an assessment system would be based on association rules identified during the construction of the graph, which would allow for a more accurate assessment of the level of understanding of the material and the identification of potential knowledge gaps. Despite the significant advantages of the proposed approach, there are certain limitations, such as the need for expert support at the stage of model setup and limited ability to automate the analysis of new topics. Future research will focus on developing more versatile algorithms that can automatically update the knowledge graph taking into account new data and trends in the blockchain field. In addition, the development of a student assessment system based on the graph will be an important step in improving the learning process.

VI. CONCLUSION

In conclusion, the conducted study shows that the application of FCA and clustering methods to create an interactive knowledge graph in educational programs on blockchain technologies is highly effective. The work developed an approach that allows teachers to flexibly adapt curricula depending on the level of students' training and their specialization, which is especially important in the context of blockchain technologies, covering a wide range of topics from cryptography to smart contract development. The interactive knowledge graph helps to systematize information on learning outcomes, topics, and key skills, which makes the course planning process more transparent and simplified. As a result of the study, it was possible to significantly reduce the teaching load by focusing on the main topics, which helps to reduce student overload and prevent them from dropping out due to difficulties. This is especially important for courses covering complex and multi-layered topics, as is the case with blockchain, where the amount of information can easily become excessive. The knowledge graph not only simplifies the selection of topics for teaching, but also helps teachers create more flexible and adaptive curricula that meet the rapidly changing requirements of the market and new technologies. It

was also found that the use of association rules and clustering methods helps to identify key relationships between topics and learning outcomes. This allows teachers and students to see a clearer picture of how various aspects of blockchain technology are interconnected and helps to better structure the learning process. The experiments demonstrated the possibility of using the knowledge graph not only for developing curricula, but also for further assessment of the level of students' knowledge. In the future, it is planned to use this tool to automatically assess the extent to which students have mastered key topics and learning outcomes, which will allow teachers to more effectively adjust educational materials and improve the educational process.

Despite the obvious advantages of the proposed approach, the study also revealed certain limitations. In particular, significant efforts are required from experts at the stage of developing the knowledge graph and setting up its structure. Further research is planned to develop algorithms that will automate the process of updating the knowledge graph and adapting its structure to new educational requirements and trends in the field of blockchain.

Thus, the proposed approach to organizing educational programs using formal conceptual analysis and clustering methods is an innovative tool that can significantly improve the effectiveness of blockchain technology training courses. The interactive knowledge graph makes the process of developing educational programs easier, more flexible and adaptive, which is especially important for such rapidly developing areas as blockchain.

ACKNOWLEDGMENT

We sincerely express our gratitude to the Ministry of Science and Higher Education of the Republic of Kazakhstan for the financial support of our work within the framework of the grant funding project № AP19679514 «A study on conceptualization of blockchain domain using text mining and formal concept analysis: focusing on teaching methodology».

REFERENCES

[1] Cunha, P. R. D., Soja, P., & Themistocleous, M. (2021). Blockchain for development: a guiding framework. Information Technology for Development, 27(3), 417–438. https://doi.org/10.1080/02681102.2021.1935453

[2] Fosso Wamba, S., Kala Kamdjoug, J. R., Epie Bawack, R., & Keogh, J. G. (2020). Bitcoin, Blockchain and Fintech: a systematic review and case studies in the supply chain. Production Planning and Control, 31(2–3), 115–142. https://doi.org/10.1080/09537287.2019.1631460

[3] Grover, P., Kar, A. K., Janssen, M., & Ilavarasan, P. V. (2019). Perceived usefulness, ease of use and user acceptance of blockchain technology for digital transactions – insights from user-generated content on Twitter. Enterprise Information Systems, 13(6), 771–800. https://doi.org/10.1080/17517575.2019.1599446

[4] Hughes, A., Park, A., Kietzmann, J., & Archer-Brown, C. (2019). Beyond Bitcoin: What blockchain and distributed ledger technologies mean for firms. Business Horizons, 62(3), 273–281. https://doi.org/10.1016/j.bushor.2019.01.002

[5] Allessie, D., Sobolewski, M., Vaccari, L., & Pignatelli, F. (2019). Blockchain for digital government (JRC115049). (Luxembourg: Publications Office of the European Union. Issue https://www.mafr.fr/media/assets/publications/blockchain-for-digital-government-2019.pdf.

[6] Galici, M., Mureddu, M., Ghiani, E., Celli, G., Pilo, F., Porcu, P., & Canetto, B. (2021). Energy blockchain for public energy communities. Applied Sciences, 11(8).

[7] Garg, P., Gupta, B., Chauhan, A. K., Sivarajah, U., Gupta, S., & Modgil, S. (2021). Measuring the perceived benefits of implementing blockchain technology in the banking sector. Technological Forecasting and Social Change, 163, Article 120407. https://doi.org/10.1016/j.techfore.2020.120407.

[8] Sharma, S. K., Dwivedi, Y. K., Misra, S. K., & Rana, N. P. (2023). Conjoint analysis of blockchain adoption challenges in government. Journal of Computer Information Systems, 1-14. https://doi.org/10.1080/08874417.2023.2185552.

[9] Kistaubayev, Y.; Mutanov, G.; Mansurova, M.; Saxenbayeva, Z.; Shakan, Y. Ethereum-Based Information System for Digital Higher Education Registry and Verification of Student Achievement Documents. Future Internet 2023, 15, 3. https://doi.org/10.3390/fi15010003.

[10] Zhaisanova, D., & Mansurova, M. (2024). Blockchain concept for the educational purposes: bibliometric analysis and conceptual structure. Procedia Computer Science, 231, 753-758. https://doi.org/https://doi.org/10.1016/j.procs.2023.12.142.

[11] Digital Business. (2023, October 18). Opyt Kazakhstana po izucheniyu blokcheyna v vuzakh masshtabiruyut na 200 universitetov v 50 stranakh mira [Kazakhstan's experience in studying blockchain in universities is being scaled to 200 universities in 50 countries]. https://digitalbusiness.kz/2023-10-18/opit-kazahstana-po-izucheniyu-blokcheyna-v-vuzah-masshtabiruyut-na-200-universitetov-v-50-stranah-mira.

[12] Digital Business. (2023, September 21). Blokchejn inzheneria i komplaens [Blockchain engineering and compliance]. https://digitalbusiness.kz/2023-09-21/blokchejn-inzheneria-i-komplaens/

[13] Hsu, Y.-C. (2024). "Mapping the Landscape of Data Science Education in Higher General Education in Taiwan: A Comprehensive Syllabi Analysis." Education Sciences 14(7).

[14] Sumangali K., Ch Aswani Kumar, Concept Lattice Simplification in Formal Concept Analysis Using Attribute Clustering. Journal of Ambient Intelligence and Humanized Computing. 2018;10(6):2327-43.

[15] Cui H, Yue G, Zou L, Liu X, Deng A. Multiple multidimensional linguistic reasoning algorithm based on property-oriented linguistic concept lattice. International Journal of Approximate Reasoning. 2021;131:80-92.

[16] W.L. Johnson, A. Valente, Tactical language and culture training systems: using artificial intelligence to teach foreign languages and cultures, in: Proceedings of the 20th National Conference on Innovative Applications of Artificial Intelligence, 2008, pp.1632–1639.

[17] J.R. van Seters, M.A. Ossevoort, J. Tramper, M. Goedhart, The influence of student characteristics on the use of adaptive e-learning material, Comput. Educ. 58 (2012) 942–952.

[18] Hao, F., J. Gao, C. Bisogni, G. Min, V. Loia and C. De Maio (2021). "Stability of three-way concepts and its application to natural language generation." Pattern Recognition Letters 149: 51-58.

[19] Lara-Bercial, P. J., M. C. Gaya-López, J.-M. Martínez-Orozco and S. Lavado-Anguera (2024). "PBL Impact on Learning Outcomes in Computer Engineering: A 12-Year Analysis." Education Sciences 14(6).

[20] Angelis, J., & Ribeiro da Silva, E. (2019). Blockchain adoption: A value driver perspective. Business Horizons, 62(3), 307–314. https://doi.org/10.1016/j.bushor.2018.12.001

[21] Garg, P., Gupta, B., Kapil, K. N., Sivarajah, U., & Gupta, S. (2023). Examining the relationship between blockchain capabilities and organizational performance in the Indian banking sector. Ann. Oper. Res. https://doi.org/10.1007/s10479-023-05254-0

[22] Hughes, A., Park, A., Kietzmann, J., & Archer-Brown, C. (2019). Beyond Bitcoin: What blockchain and distributed ledger technologies mean for firms. Business Horizons, 62 (3), 273–281. https://doi.org/10.1016/j.bushor.2019.01.002.

[23] Scholl, H. J., & Bolívar, M. P. R. (2019). Regulation as both enabler of technology use and global competitive tool: The Gibraltar case. Government Information Quarterly, 36(3), 601–613. https://doi.org/10.1016/j.giq.2019.05.003

[24] Cowart, J. and Y. Jin (2024). "Leading Online Professional Development for Instructional Technology Coaches with Effective Design Elements." Education Sciences 14(7).

[25] Chang, M., G. D'Aniello, M. Gaeta, F. Orciuoli, D. Sampson and C. Simonelli (2020). "Building Ontology-Driven Tutoring Models for Intelligent Tutoring Systems Using Data Mining." IEEE Access 8: 48151-48162.

[26] Cristea, D., D. Șotropa, A.-J. Molnar and S. Motogna (2021). On the Use of FCA Models in Static Analysis Tools to Detect Common Errors in Programming. Graph-Based Representation and Reasoning: 3-18.

[27] Hao F, Gao J, Bisogni C, Min G, Loia V, De Maio C. Stability of three-way concepts and its application to natural language generation. Pattern Recognition Letters. 2021;149:51-8.

[28] Muangprathub, J., V. Boonjing and K. Chamnongthai (2020). "Learning recommendation with formal concept analysis for intelligent tutoring system." Heliyon 6(10): e05227.

[29] Santana, F., Almeida, J. P. A., & Guizzardi, G. (2018). Ontology engineering for improving interoperability, semantic integration, and knowledge discovery in cryptocurrency platforms. Journal of Web Semantics, 52, 29-44. https://doi.org/10.1016/j.websem.2018.07.004

[30] Song H, Zhu N, Xue R, He J, Zhang K, Wang J. Proof-of-Contribution consensus mechanism for blockchain and its application in intellectual property protection. Information Processing & Management. 2021;58(3).

[31] Betarte, G., Cristiá, M., Luna, C., Silveira, A., Zanarini, D. (2020). Towards a Formally Verified Implementation of the MimbleWimble Cryptocurrency Protocol. In: Zhou, J., et al. Applied Cryptography and Network Security Workshops. ACNS 2020. Lecture Notes in Computer Science(), vol 12418. Springer, Cham. https://doi.org/10.1007/978-3-030-61638-0_1

[32] Son D-H. On-demand ride-sourcing markets with cryptocurrency-based fare-reward scheme. Transportation Research Part E: Logistics and Transportation Review. 2023;171.

[33] Kobayakawa N, Imamura M, Nakagawa K, Yoshida K. Impact of Cryptocurrency Market Capitalization on Open Source Software Participation. Journal of Information Processing. 2020;28(0):650-7.

[34] Vidal-Tomás D. The new crypto niche: NFTs, play-to-earn, and metaverse tokens. Finance Research Letters. 2022;47.

[35] Aquilina M, Frost J, Schrimpf A. Tackling the risks in crypto: Choosing among bans, containment and regulation. Journal of the Japanese and International Economies. 2024;71.

[36] Subramanian H, Rouxelin F. Do cryptocurrency rewards improve platform valuations? Information & Management. 2024;61(6).

[37] Hajiaghapour-Moghimi M, Hajipour E, Azimi Hosseini K, Vakilian M, Lehtonen M. Cryptocurrency mining as a novel virtual energy storage system in islanded and grid-connected microgrids. International Journal of Electrical Power & Energy Systems. 2024;158.

[38] Hajdu, D., Benczur, A., & Kertész, J. (2020). Formal Concept Analysis in the realm of cryptocurrency: Uncovering contextual nuances and semantic relationships. Journal of Cryptographic Research, 35(4), 212-228. https://doi.org/10.1016/j.jcr.2020.04.005

[39] Kölbl, T., Müller, R., & Schmid, F. (2019). Context-aware ontology construction using Formal Concept Analysis in the cryptocurrency ecosystem. International Journal of Semantic Computing, 13(2), 155-178. https://doi.org/10.1142/S1793351X19400042

[40] Ganter, B., & Wille, R. (1999). Formal Concept Analysis: Mathematical Foundations. Springer.

Analysis of Influencing Factors of Tourist Attractions Accessibility Based on Machine Learning Algorithm

Na Liu*, Hai Zhang
School of Traffic and Transportation, Xi'an Traffic Engineering Institute, Xi'an, Shaanxi, 710300, China

***Abstract*—Tourist attractions, defined by their cultural importance, aesthetic appeal, and recreational possibilities, are critical to the tourism industry. However, precisely evaluating tourism needs remains a difficult task, and research in this field is scarce. This research introduces an innovative remora-optimized adaptive XGBoost (RO-AXGBoost) model for predicting accessibility factors for tourist attractions. Data was obtained from Kaggle, and the suggested method was executed in Python. The RO-AXGBoost model's effectiveness was assessed utilizing metrics like Mean Absolute Percentage Error (MAPE) of 7.24, Mean Absolute Error (MAE) of 7.321, Root Mean Square Error (RMSE) of 10.241, and R-squared (R^2) of 85.7%. The results show that the RO-AXGBoost model surpasses conventional approaches by effectively discovering important determinants that have an important impact on the accessibility of tourist attractions.**

Keywords*—Tourist attractions; factors; tourism; remora optimized adaptive XGBoost (RO-AXGBoost)

I. Introduction

The intention to return is a vital component of tourism study, especially in the setting of urban tourism, where tourist loyalty is critical to economic development [1]. Comprehending the factors that impact tourists' choices to revisit attractions not only helps with consumer retention but also improves the efficacy of advertising tactics. With the advent of social media, tourists now have unprecedented accessibility to data, which greatly influences their travel decisions. Social media is an effective tool for influencing thoughts and experiences, establishing itself as an indispensable resource in the vacation decision-making procedure [2].

As social media platforms spread, the rise of social media influencers (SMIs) has altered how data is distributed [3]. These individuals, operating as independent advocates, can influence public opinion via a variety of digital channels, including tweets and blog posts. This dynamic shift in data access is especially important in the tourism industry, which is naturally vulnerable to external shocks like epidemics and catastrophes. The World Health Organization (WHO) has frequently suggested halting economic operations and public events during emergencies, demonstrating the industry's vulnerabilities.

Furthermore, environmental issues are posing new challenges to the modern tourism industry. As travelers become more conscious of the environmental effects of their operations, the need for "sustainable tourists"—those with smaller environmental impacts and more buying power—has increased [4]. The tourism industry is thus forced to adjust to these changing customer desires by incorporating environmentally friendly procedures.

Technological developments have become integral to the growth of the tourism industry, resulting in the idea of smart tourism [5]. This digital transformation not only improves tourism resource management but also provides stakeholders with timely access to pertinent data. The incorporation of sophisticated information technologies enables connection between tourism providers, thus improving the entire experience of visitors.

However, an important obstacle remains guaranteeing the safety and security of travel destinations. The growing issues of civil unrest, terrorism, catastrophic events, and health emergencies have increased the demand for dependable security evaluations [6]. Tackling these problems is essential to efficient travel planning.

In this setting, the purpose of this research is to improve the accuracy of predicting tourist attraction variables using an innovative remora-optimized adaptive XGBoost (RO-AXGBoost) model. This sophisticated machine learning method is expected to enhance predictive abilities in terms of important factors impacting tourist destinations. This study not only adds to the body of knowledge in tourism management but also provides decision-makers with useful information for tactical planning.

Additionally, to these advances in tourism management, current developments in image processing methods have substantially enhanced the excellence of visual content related to tourist attractions. Deep learning-based image improvement methods and the use of artificial intelligence algorithms have developed as methods for removing haze from images, leading to more obvious and appealing visuals. These improvements not only improve the visual appeal of tourist destinations, but they additionally perform an important role in impacting tourists' thoughts, eventually impacting their travel choices.

The accessibility of tourist attractions is critical to the achievement and enjoyment of tourism experiences. However, precisely predicting and evaluating the variables that impact accessibility is a difficult and underexplored problem. Conventional techniques frequently fall short of capturing the complex interplay of factors like transportation infrastructure, ecological sustainability, and tourist behaviour. The growth of smart technologies, as well as varying external conditions such as epidemics and climate change, have intricate the tourism sector even further. This study aims to close the gap by using sophisticated machine learning methods to present more precise, scalable, and flexible predictions of the variables that influence the accessibility of tourist attractions.

*Corresponding Author.

This study intends to answer the following important questions:

1) What are the most crucial elements impacting the accessibility of tourist attractions?

2) How can machine learning techniques, especially the RO-AXGBoost model, be used to enhance the prediction accuracy of tourist destination accessibility?

3) How does the suggested RO-AXGBoost compare with other previous methods in terms of performance metrics like MAPE, MAE, RMSE, and R^2?

4) What potential knowledge can be derived from utilizing sophisticated predictive models for tourism management and planning?

The primary objectives of this study are:

1) To detect and evaluate the important determinants impacting the accessibility of tourist attractions.

2) To create and execute a new remora-optimized adaptive XGBoost (RO-AXGBoost) model that enhances the prediction accuracy of accessibility variables.

3) To compare the suggested method with previous machine learning techniques and assess its effectiveness utilizing metrics like MAPE, MAE, RMSE, and R^2.

4) To present knowledge and suggestions for tourism managers and urban planners on how to improve accessibility and sustainability in tourist destinations.

The paper's structure aims to direct the reader through the procedure of research, beginning with an introduction that presents background data on the significance of accessibility in tourism as well as technological advances that have allowed for more accurate prediction models in Section I. This is followed by a thorough literature review in Section II, which investigates previous research on machine learning methods in tourism and identifies gaps in current methodologies. The methodology in Section III describes the data sources, model creation, and performance assessment criteria, whereas the outcomes section summarizes the research's findings and compares the suggested RO-AXGBoost to previous techniques. The discussion in Section IV examines the results critically, emphasizing the study's novelty and implications. Finally, the paper summarizes the important contributions and proposes future research directions in Section V to build on the findings of this research.

II. Related Work

A study in [7] examined seasonal emotional shifts among travelers in natural forest environments using advanced methods such as Hrnet, SHAP, OSANetand, and XGBoost. Positive feelings were most common throughout three years, with varying distributions among seasons. The study contributed to the development of sustainable forest tourism by revealing subtle emotional reactions to landscape aspects.

Research in [8] used an IoT-enabled, deep learning-based DNN with a multi-class classification algorithm recommendation system to improve visitor experiences in smart cities. Personalized suggestions were informed by real-time data and feedback from travelers. Compared to previous models, our multi-label classifier performed better, attaining high accuracy and precision. Potential biases in the source data and the requirement for a substantial IoT infrastructure were examples of limitations.

Paper in [9] employed an improved STC-LSTM to forecast short-term vacation demand for travel. As compared to traditional approaches, the results show enhanced forecast accuracy. The limitations of the approach encompass dependence on certain data categories and possible problems with the model's regional generalization.

Research in [10] proposed a unique strategy centered on enhancing visitor experiences and reducing congestion to overcome the shortcomings of conventional T-RSs. Used deep reinforcement learning as opposed to just single areas of interest. Its effectiveness was evaluated against three baselines using Verona tourism data from 2014 to 2023. The heavy computing load and data requirements were among the drawbacks.

Research in study [11] proposed a three-stage methodology to anticipate tourism demand across numerous sites. It utilized multi-dimensional scaling to identify related attractions, mixed autoregressive models with LSTM networks to reflect spatial dependency and the scale of tourism and suggested a method for integrating predictors to increase prediction stability. The program performed exceptionally well in estimating the number of visitors to Beijing's 77 attractions, according to the results. Data accessibility and generalizability issues were two possible limitations.

The objective of the research [12] was to create a DNN model for predicting how well policies would function over time to lessen the effects of crises on the travel and tourism sector in developing nations. The model indicated that the most efficient and long-lasting strategy was to concentrate on both domestic tourist marketing and disaster preparedness, drawing on both past experiences and present difficulties. The intricacy of modeling dynamic systems and the availability of data were possible constraints.

A study in [13], which focused on incoming arrivals from Hong Kong, attempted to improve the accuracy of tourist volume predictions using deep learning techniques. The suggested SAE-Bi-GRU strategy beat benchmark models like PCA-Bi-GRU with the Baidu index and Google trends data. It was archived by using SAE for data dimension reduction and a Bi-GRU for model forecasting. The research's dependence on a single case study, however, presented a limitation.

Study [14] improved tourist analysis by combining deep-learning text categorization and spatial clustering approaches. Using Flickr data, it created nine tourist categories using topic modeling and LSTM classification. Spatial clustering determined the ROA for each category, indicating various elements impacting attraction. The technique provided extensive insights into visitor tastes and could be used outside tourism, although its dependence on Flickr data has limits.

The purpose of the study [15] was to better understand travelers' binary emotional experiences at Dali tourist spots. SVM and LDA models were used, as well as geographic analysis. The results reflect a general trend of good attitude,

with negative evaluations for select Dali city attractions. Service experience and pricing were revealed as important elements affecting travelers' attitudes. Limitations may include a dependence on internet evaluations and the study's location.

III. METHODOLOGY

In this section, the paper proposed a novel remora-optimized adaptive XGBoost (RO-AXGBoost) to forecast various factors associated with tourist attractions. We gathered data from Kaggle.

A. Data Collection

This paper gathers a dataset from Kaggle [16]; this dataset covers tourism attraction site data from 352 Chinese cities. Each city CSV file has 100 locations. The data comprises the location name, URL, address, site introduction, opening hours, image URL, rating, proposed visit duration, suggested season, ticket information, and tips.

B. Remora Optimized Adaptive XGBoost

The suggested technique, remora-optimized adaptive XGBoost (RO-AXGBoost), was selected for its capacity to manage intricate datasets and make precise predictions about accessibility elements of tourist attractions. This technique stands out because it incorporates optimization methods that improve the model's efficiency, rendering it especially appropriate for evaluating the variety of tourism data. This research's participants are a varied group of tourists with varying demographics like age, gender, nationality, and travel experience, guaranteeing a thorough comprehension of various viewpoints on accessibility. The study will make use of publicly available datasets, like those obtained from Kaggle, which include variables like travel frequency, desired attractions, accessibility issues, and socioeconomic status. This integration of sophisticated machine learning methods and large datasets intends to offer resilient knowledge of the factors that influence the usability of tourist attractions, eventually leading to better tourism management procedures.

RO-AXGBoost is a novel technique for anticipating several aspects associated with tourist attractions. It leverages the Remora Optimization technique to adjust XGboost parameters and operates at the best model optimization. By coordinating Remora's flexibility with XGBoost's robustness, RO-AXGBoost is a potent tool for tourist destination management and planning as it enhances the machine's accuracy and efficiency in predicting crucial measurable factors in the tourism sector industry. The integration of Remora's suppleness with the robustness of XGBoost is called RO-AXGBoost, which presents viable aid to key decision-makers within the tourism sector that can forecast the significant indicators necessary for strategic planning and management of destination tourism more effectively.

1) ROA: This section aims at introducing the new meta-heuristic that has been developed, known as the ROA. ROA is based on a symbiotic float trait that remoras use to cling to bigger living organisms like swordfish and whales to be in a better position to access food. Again, like all of the MAs, the ROA optimization technique's concept is derived from biological considerations. This one has incorporated some of the features from the WOA and the SCA, which is known for its efficiency in both the local and the global search. By selecting an integer variable called H (0 or 1), one can decide whether this strategy has to use a similar exploring manner to that in SCA or WOA. This flexible manner increases using search space in the case of ROA, but if compared to other approaches to refinement, this flexible manner decreases the level of accuracy in exploration.

a) Travel with free: Using an elite approach, the SFO strategy is employed by ROA to perform a global search in the swordfish algorithm. The Eq. (1) for updating positions can be stated as follows:

$$u_i(p+1) = X_{best}(p) - (rand \times \left(\frac{x_{best}(p)+x_{rand}(p)}{2}\right) - x_{rand}(p)) \quad (1)$$

In this instance $\boldsymbol{u_i(p+1)}$ represents a candidate position of the $i-th$ remora. At the present moment, the optimal position is$\boldsymbol{X_{best}}(\boldsymbol{p})$. An arbitrary remora location is denoted by$\boldsymbol{x_{rand}}(\boldsymbol{p})$. The letter$P$stands for the number of iterations that are currently in progress. And between 0 and 1, $rand$ is a random number. Additionally, depending on its experiences, remora may switch hosts. This situation allows for the creation of a new candidate position through:

$$u_i'(p+1) = u_i(p+1) + randn \times \left(u_i(t+1) - x_i(t)\right) \quad (2)$$

And the $i-th$ remora's candidate position is represented by $u_i'(p+1)$. The ith remora's prior location is denoted by $x_i(t)$. Additionally, a properly distributed random number is generated using $randn$.

b) Through fully eat: Remora may also attach themselves to humpback whales to feed. Therefore, remora will move similarly to humpback whales. The WOA method is used in ROA to do local searches. To be more exact, the bubble-net attacking approach employed in WOA is utilized. The changed position updating formulae are provided below:

$$u_i(p+1) = E \times f^S \times \cos(2\pi a) + x_{best}(p) \quad (3)$$

$$E = |X_{best}(p) - x_i(p)| \quad (4)$$

Here E is the separation between the food and the remora Eq. (4). Additionally, by employing the encircling prey mechanism in WOA, which is described as follows, the remora can generate a little step to further enhance the quality of the solution.

$$E' = u_i(p+1) - S \times x_{best}(p) \quad (5)$$

$$x_i(p+1) = v_i(p+1) + B \times E' \quad (6)$$

Here the newly created location of the $i-th$ remora is denoted by $u_i(p+1)$. In ROA, the remora factor, represented by the letter B, is set to 0.1. When using the aforementioned techniques, ROA outperforms well-known meta-heuristic algorithms like WOA, SFO, and HHO (see Eq. (5) and (6)).

2) Adaptive extreme Gradient Boosting (AXGBoost): In this study, we employ AXGBoost as the core method to forecast gains and handle high-dimensional assembly faults. Then, to

get predictions that are even more accurate, we suggest adaptiveXGBoost.

$$\hat{o}_i = h_2(E_i) = \sum_{k_2=1}^{k_2} h_{k_2}(E_i), h_{k_2} \in \emptyset \quad (7)$$

Here $\phi = \{h(E) = \omega_{s(D)}\}$ is a grouping of decision trees. Each $tree\ h(E)$ corresponds to a structural parameter s and leaf weights ω, Eq. (7). The $i-th$analysis of tourist attractions is denoted by$\hat{o}_i$. The model is developed by minimizing the total loss function:

$$M = \sum_{i=1}^{N_t} m(\hat{o}_i, o_i) + \sum_{k_2=1}^{k_2} \Omega(h_{k_2}), \quad (8)$$

Here,

$$\Omega(h_{k_2}) = \gamma d + \varepsilon \|\omega\|^2 \quad (9)$$

$m(\hat{o}_i, o_i)$Refers to the loss function used to calculate the difference between the realistic and expected tourism attraction. dRepresents the number of decision tree leaves. The penalty period is represented by the symbol Ω, Eq. (8) and (9). The tuning parameters γ and ε regulate the complexity of decision trees. The square loss function was employed in this investigation.

$$m(\hat{o}_i, o_i) = (\hat{o}_i, -o_i)^2 \quad (10)$$

Iterative training is used for the loss function. For the $s-th$ repetition of the $i-th$ sample, we might write Eq. (8) as:

$$M^d = \sum_{i=1}^{N_t} m(\hat{o}_i^{(d-1)} + h_d(E_i), o_i) + \Omega(h_s) \quad (11)$$

XGBoost enhances the model by addingh_s, Eq. (10). The objective is optimized using a second-order approximation.

$$M^d \approx \sum_{i=1}^{N_t} \left[m(\hat{o}_i^{(t-1)}, o_i) + t_i h_i(E_i) + \frac{1}{2} p_i h_s^2(E_i)\right] + \Omega\ (h_s) \quad (12)$$

$$T_i = \varphi_{\hat{o}_i^{(t-1)}} m(\hat{o}_i^{(s-1)}, o_i) = 2 \times (\hat{o}_i^{(s-1)} - o_i), \quad (13)$$

$$G_i = \varphi_{\hat{o}_i^{(s-1)}} \ ^2m\left(\varphi_{\hat{o}_i^{(s-1)}}, o_i\right) = 2 \quad (14)$$

The first and second-order gradients are denoted as $T_i$and G_i. The equations are$m\left(\varphi_{\hat{o}_i^{(s-1)}}, o_i\right)$ Eq. (12)-(14).

An adaptive XGBoost model serves as the foundation for transfer learning, which uses parameter-based approaches. During the process of the transfer training phase, leaf weights from all trees in the adaptive XGBoost model are exchanged. It is constructed by training new trees and leaf weights. Fig. 1 illustrates the adaptive XGBoost framework.

$$\hat{o}_{2,i} = g_3(E_i) = \sum_{K_2=1}^{K_2} h_{k_2}(E_i) + \sum_{K_5=1}^{K_5} h_{k_2}(E_i), h_{k_2,} h_{k_5,} \quad (15)$$

The TF model predicts tourism for the $i-th$ sample, denoted as$\hat{o}_{2,i}$. $h_{k_2,}$Represents the trees in the enhanced XGBoost model, whereas $h_{k_5,}$is in Eq. (15). To increase the accuracy of tourism attraction predictions, the weights of leaves in $\boldsymbol{h_{k_2,}}$ are shared and newer leaf strengths in $\boldsymbol{h_{k_5,}}$are trained.

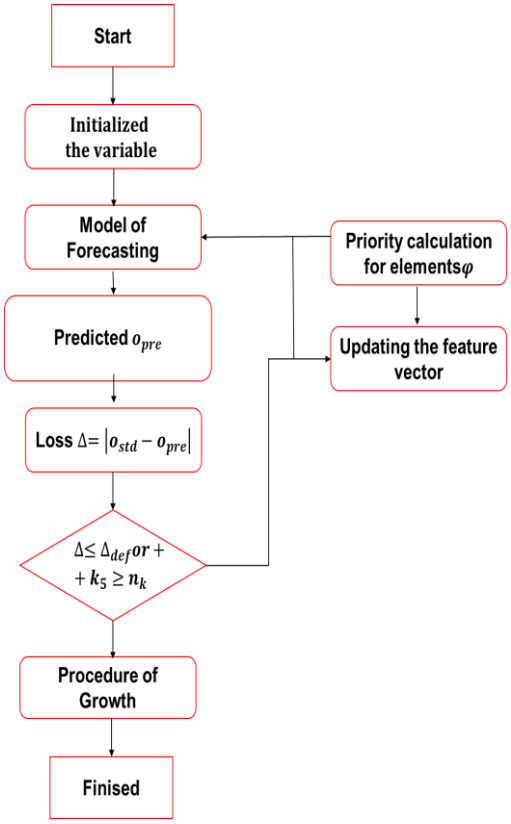


Fig. 1. Adaptive XGBoost.

3) Predicting RO-AXGBoost: The original technique that we used is called RO-AXGBoost, and it incorporates Remora Optimization, a newly developed meta-heuristic algorithm that refers to the behavior of remoras about sharks in improving the capabilities of XGBoost in predicting various aspects that are related to tourist attractions. Thus, our method of utilizing the Remora Optimization procedure to enhance the adaptiveness of XGBoost hyper-parameters enhances its quality in terms of its performance and capabilities of precisely predicting labeled tourism-related variables. These two components allow for better and more consistent forecasting, helping the entities in the tourism sector to optimize their decision-making process in terms of resource management, promotion, and visitors' handling, thereby contributing to the overall improvement of the visitors' experience and the sustainability of the destinations.

IV. Result and Discussion

Anaconda3 for Windows is utilized in this scenario on a PC powered by an NVIDIA GeForce RTX 3090 Ti GPU and 24GB of RAM. The hard drive capacity is 1TB, and the computer's processor is the Intel Core i7-8750H running at 2.20GHz. Python 3.10 is the main programming language, using TensorFlow 1.14.0 as a development framework. In the section evaluate the performance outcome for the proposed method. The proposed method is compared with existing methods and the performance is evaluated in different parameters such as MAPE, MAE, RMSE, and R^2. The existing methods are

BiLSTM [18], CNN-BiLSTM [18], MACBL [18], KNN [17], RF [17] and LR [17].

R^2 is an indicator of how accurately the variables that are independent predict the dependent variable. Fig. 2 illustrates on R^2 value. Compared to existing methods KNN (70.6), RF (74.9), and LR (76.5), our proposed method was superior (RO-AXGB 85.7). In comparison to existing methods, the suggested method RO-AXGB showed significant improvements in tourist attraction accessibility.

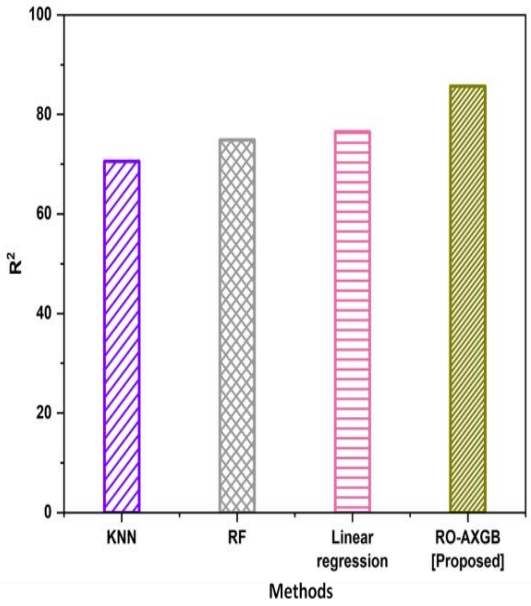


Fig. 2. Outcome of R^2.

The RMSE is the average of the squared discrepancies between anticipated and actual values. It indicates how much the model's predictions differ from the actual results, with lower values representing better performance. Fig. 3 illustrates the RMSE result. In comparison to the existing techniques BiLSTM (23.326), CNN-BiLSTM (16.354), and MACBL (13.837), our proposed is (RO-AXGB -10.241) lower than existing methods. It shows that our suggested method, RO-AXGB is effective for tourist attractions accessibility.

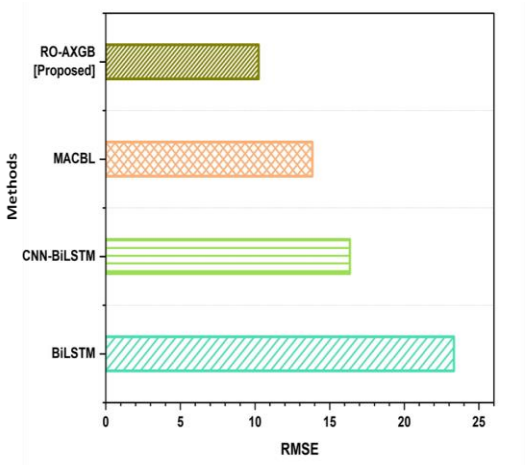


Fig. 3. Outcome of RMSE.

MAE estimates the average of the absolute differences between anticipated and actual values. It reveals information about the model's accuracy in forecasting, but it is less sensitive to extremes since it does not square the errors. Fig. 4 illustrates the MAE values. When compared to existing approaches BiLSTM (19.146), CNN-BiLSTM (11.354), and MACBL (8.928), our suggested technique is higher (RO-AXGB -7.321) than the existing method. It demonstrates that our suggested approach, RO-AXGB, successfully detects tourist attractions accessibility.

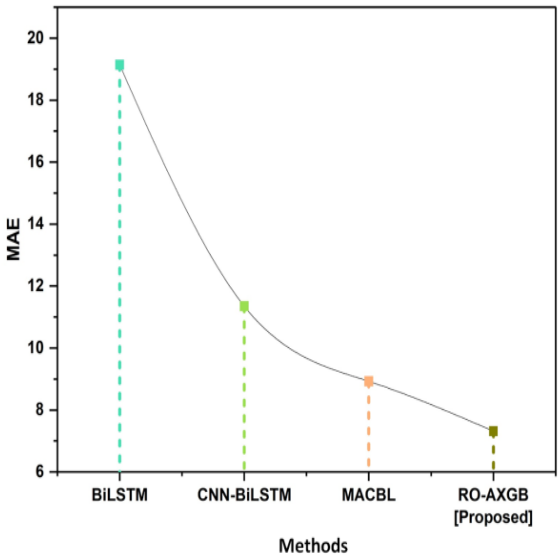


Fig. 4. Outcome of MAE.

MAPE is the average of the absolute percentage deviations between projected and actual values, represented as a percentage of actual values. Fig. 5 depicts the MAPE result and Table I illustrates the comparison of existing. When compared to the existing methods, BiLSTM (21.42), CNN-BiLSTM(16.96), and MACBL (10.21), our suggested method performs better (RO-AXGB -7.24). The successful demonstration of a tourist attraction's accessibility using the RO-AXGB approach indicates its effectiveness.

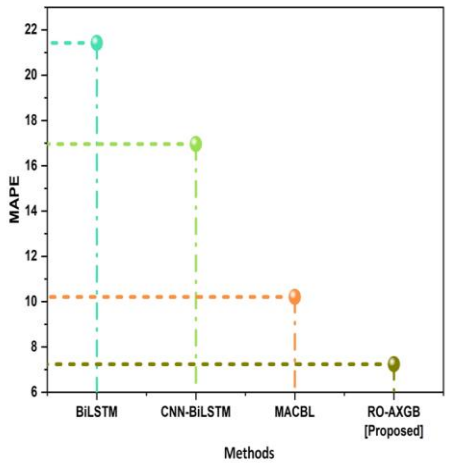


Fig. 5. Outcome of MAPE.

TABLE I. COMPARISON OF EXISTING METHODS WITH PROPOSED METHOD

Method	RMSE	MAE	MAPE	Method	R^2
BiLSTM	23.326	19.146	21.42	KNN	70.6
CNN-BiLSTM	16.354	11.354	16.96	RF	74.9
MACBL	13.837	8.928	10.21	LR	76.5
RO-AXGB	10.241	7.321	7.24	RO-AXGB	85.7

A. Discussion

Though the current methods of classifying influential factors of tourist attractions' accessibilities as BiLSTM, CNN-BiLSTM, MACBL, KNN, RF, and LR shed light on the circumstances of the objective, they all have their drawbacks. For that reason, both BiLSTM [18] and CNN-BiLSTM [18] models perform well in modeling sequential dependencies as well as spatial relations but can fail to adequately model long-range dependency and global context information. To address this issue, MACBL [18] employs a multiple-attentional mechanism in the feature extraction area to improve performance, but the additional attention mechanisms might increase the number of parameters for the resulting model and risk overfitting.

KNN [17], although the highlighted algorithm is easy to understand and implement, has its limitations regarding its scalability and effectiveness when applied to datasets containing high dimensions. RF [17] has high resistance to overfitting and nonlinearity but can have a low capacity for modeling high-level interactions among the variables. As a flexible, interpretable, and easy-to-implement method, LR [17] might not have the ability to capture the interactions and complex functional forms between the variables. Thus, although all of these methods yield valuable insights to analyze tourist attractions' accessibility, their flawed aspects point out to continued need for research and possibly a combination of all the methods to provide effective solutions regarding the various angles of the problem.

The drawback tackled, in this paper RO-AXGBoost provides an innovative approach for the interpretation of perturbing variables in the case of tourist attraction availability. Effectively applying the optimization techniques and the incorporation of adaptability to the equations make RO-AXGBoost highly efficient in capturing patterns of accessibility, which can aid in sound decision-making processes in the tourism sector. Its strength can be attributed to its capability to deal with big data, particularly concerning operating and learning in fluctuating environments and delivering solutions with noteworthy speed and accuracy.

This study's results show that the suggested remora-optimized adaptive XGBoost (RO-AXGBoost) technique outperforms previous methods in forecasting accessibility factors for tourist attractions. RO-AXGBoost outperformed other techniques (KNN, RF, and LR) in predicting accessibility-related variables, with an R^2 value of 85.7%. Additionally, the technique surpassed models like BiLSTM, CNN-BiLSTM, and MACBL in terms of MAE (7.321) and MAPE (7.24), demonstrating its superior accuracy in reducing prediction errors. However, in terms of RMSE, RO-AXGBoost produced a moderate outcome, indicating the necessity for future improvement in some situations. Compared to previous research that used BiLSTM and CNN-BiLSTM, which are efficient at capturing sequential data but have constrained long-range dependency managing, RO-AXGBoost's incorporation of optimization methods and adaptive learning proved useful in handling larger datasets and intricate interactions. This study presents a new combination of remora optimization and XGBoost, a method that not only improves efficiency but also tackles the constraints of previous models, resulting in more precise and effective predictions in tourism accessibility research.

To guarantee the resilience and validity of the suggested RO-AXGBoost method, cross-validation methods like k-fold cross-validation were used to decrease possible overfitting while also providing an accurate assessment of the model's generalization capacity. Furthermore, the suggested approach was compared to closely associated methods, like Gradient Boosting Machines (GBM) and XGBoost, which use boosting methods to enhance efficiency. While both GBM and XGBoost provide significant accuracy gains in intricate datasets, RO-AXGBoost incorporates adaptive optimization and feature weighting tactics, improving its capacity to handle high-dimensional data more efficiently. In comparison to XGBoost, which can struggle with particular dynamic trends, the RO-AXGBoost method handled accessibility-related perturbations more effectively because of its flexibility in handling various variables. Additionally, unlike GBM, which has higher computational expenses, RO-AXGBoost optimized execution speed, especially in environments demanding rapid decisions, rendering it ideal for tourism applications where quick accessibility knowledge are critical. The findings show that the RO-AXGBoost method improves both accuracy and speed, surpassing comparable techniques.

V. CONCLUSION

This research proposed an innovative remora-optimized adaptive XGBoost (RO-AXGBoost) technique for predicting important parameters influencing tourist attraction accessibility. The technique outperformed previous algorithms in prediction accuracy and precision, as measured by famous metrics like MAPE, MAE, RMSE, and R^2. Key factors impacting accessibility were successfully discovered, presenting useful information for tourism management and urban planning. However, while the technique produced impressive outcomes, some constraints must be acknowledged. The moderate RMSE values indicate possible fields for improvement, especially in dealing with variability and outliers in the data. Furthermore, while the study concentrated mainly on accessibility, further investigation into other dimensions of tourist satisfaction, like convenience and environmental sustainability, could offer a more extensive picture of the tourism industry's difficulties.

Future work could include incorporating real-time data to improve the model's flexibility in dynamic settings, as well as investigating more sophisticated optimization methods to further decrease prediction errors. Furthermore, integrating factors such as tourist desires, environmental concerns, and

sustainability into the model would provide more detailed insights into tourism planning and management. Extending the model to include various kinds of attractions and geographies may also enhance its generalizability and applicability in wider tourism settings.

REFERENCES

[1] Hussein, A.S., "City branding and urban tourist revisit intention: The mediation role of city image and Visitor satisfaction", International *Journal of Tourism Policy*, vol. 10, no. 3, pp.262-279, 2020.

[2] Rather, R.A., "Monitoring the impacts of tourism-based social media, risk perception, and fear on tourist's attitudes and revisiting behavior in the wake of the COVID-19 pandemic", Current *Issues in Tourism*, vol. 24, no. 23, pp.3275-3283, 2021.

[3] Pop, R.A., Săplăcan, Z., Dabija, D.C. and Alt, M.A., "The impact of social media influencers on travel decisions: The role of trust in consumer decision journey", *Current Issues in Tourism*, vol. 25, no. 5, pp.823-843, 2022.

[4] Holmes, M.R., Dodds, R. and Frochot, I., "At home or abroad, does our behavior change? Examining how everyday behavior influences sustainable travel behavior and tourist clusters" *Journal of Travel Research*, vol. 60, no. 1, pp.102-116, 2021.

[5] Azis, N., Amin, M., Chan, S. and Aprilia, C., "How smart tourism technologies affect tourist destination loyalty", *Journal of Hospitality and Tourism Technology*, vol. 11, no. 4, pp.603-625, 2020.

[6] Sharma, A.K., Birendra, K.C. and Calderon, A.H., "Perceived safety and security concerns among tourists in Thamel-a tourism hub in Kathmandu Valley, Nepal", *Indonesian Journal of Tourism and Leisure*, vol. 1, no. 2, pp.92-102, 2020.

[7] Chen, Z., Ye, C., Yang, H., Ye, P., Xie, Y. and Ding, Z., "Exploring the impact of seasonal forest landscapes on tourist emotions using Machine learning", *Ecological Indicators*, p.112115, 2024.

[8] Cepeda-Pacheco, J.C. and Domingo, M.C., "Deep learning and Internet of Things for tourist attraction recommendations in smart cities", *Neural Computing and Applications*, vol. 34, no. 10, pp.7691-7709, 2022.

[9] Li, W., Guan, H., Han, Y., Zhu, H. and Wang, A., "Short-term holiday travel demand prediction for urban tour transportation: a combined model based on STC-LSTM deep learning approach", *KSCE Journal of Civil Engineering*, vol. 26, no. 9, pp.4086-4102, 2022.

[10] DallaVecchia, A., Migliorini, S., Quintarelli, E., Gambini, M. and Belussi, A., "Promoting sustainable tourism by recommending sequences of attractions with deep reinforcement learning", *Information Technology & Tourism*, pp.1-36, 2024.

[11] Bi, J.W., Han, T.Y. and Yao, Y., "Collaborative forecasting of tourism demand for multiple tourist attractions with spatial dependence: A combined deep learning model" *Tourism Economics*, vol. 30, no. 2, pp.361-388, 2024.

[12] Neshat, N., Moayedfar, S., Rezaee, K. and AmrollahiBiuki, N., "Sustainable planning of developing tourism destinations after COVID-19 outbreak: A deep learning approach", *Journal of Policy Research in Tourism, Leisure and Events*, vol. 16, no. 1, pp.1-21, 2024.

[13] Li, M., Zhang, C., Sun, S., and Wang, S., "A novel deep learning approach for tourism volume forecasting with tourist search data", *International Journal of Tourism Research*, vol. 25, no. 2, pp.183-197, 2023.

[14] Lee, H. and Kang, Y., "Mining tourists' destinations and preferences through LSTM-based text classification and spatial clustering using Flickr data", *Spatial Information Research*, vol. 29, no. 6, pp.825-839, 2021.

[15] Yin, X. and Jung, T., "Analyzing the causes of tourists' emotional experience related to tourist attractions from a binary emotions perspective utilizing machine learning models", *Asia Pacific Journal of Tourism Research*, pp.1-20, 2024.

[16] The data is available online: https://www.kaggle.com/datasets/audreyhengruizhang/china-city-attraction-details?resource=download

[17] Bravo, J., Alarcón, R., Valdivia, C. and Serquén, O., "Application of Machine Learning Techniques to Predict Visitors to the Tourist Attractions of the Moche Route in Peru", *Sustainability*, vol. 15, no. 11, p.8967, 2023.

[18] Tang, Q., Yang, L. and Pan, L., "Passenger flow forecast of tourist attraction based on MACBL in LBS big data environment", *Open Geosciences*, vol. 15, no. 1, p.20220577, 2023.

A Smart Contract Approach for Efficient Transportation Management

Abdullah Alshahrani[1], Ayman Khedr[2], Mohamed Belal[3], Mohamed Saleh[4]
University of Jeddah, College of Computer Science and Engineering,
Department of Computer Science and Artificial Intelligence, Jeddah, Saudi Arabia[1]
University of Jeddah, College of Computing and Information Technology,
Department of Information Systems, Jeddah, Saudi Arabia[2]
Faculty of Computers and Artificial Intelligence, Department of Computer Science, Helwan University, Cairo, Egypt[3]
Faculty of Commerce and Business Administration, Department of Business Information Systems,
Helwan University, Cairo, Egypt[4]

***Abstract*—Transportation management in Egypt faces challenges such as congestion, inefficiency, and a lack of transparency. This work proposes a smart contract-based transportation framework to address these issues and enhance the efficiency of Egypt's transportation system. By leveraging blockchain technology, smart contracts can facilitate and enforce decentralized and immutable transportation agreements. This approach also fosters increased trust among stakeholders and improves interactions between service providers. This paper presents a conceptual framework that integrates smart contracts, blockchain technology, GPS data, and sensor technologies to further optimize transportation operations. Empirical analysis and case studies demonstrate the effectiveness of smart contracts in improving the shipping registration system. The survey results show that smart contracts streamline processes enhance data security, reduce costs, and improve accuracy. The proposed model, developed on the NEAR platform, outperforms traditional methods and Ethereum-based models by offering faster registration, better cost-efficiency, and improved transaction tracking. This demonstrates the potential for modernizing and optimizing Egypt's transportation sector**.

***Keywords*—Blockchain; cryptography; logistics; smart contracts; transportation; security and privacy; supply chains**

I. Introduction

This Transportation is vital to the global economy, as it connects resources for manufacturing and delivers finished products to customers. However, this fundamental industry faces challenges such as speed, cost-efficiency, and safety [1]. The process of transportation is intricate and demands careful planning, coordination, and implementation. The transportation process commences with route planning and the selection of carriers [2]. Companies must determine the most efficient and cost-effective routes while selecting the appropriate transportation carriers based on various factors. This decision impacts both the timeliness of delivery and the cost-effectiveness of the entire process. Freight booking, loading, unloading, and proper documentation ensure the security, tracking, and legal compliance of goods throughout their journey [1]. Freight tracking, last-mile delivery, risk management, effective communication, and coordination are essential components that keep the supply chain running smoothly [3]. Modern technology, including GPS and communication systems, has revolutionized transportation by enabling real-time tracking and responsive problem-solving. In the end, transportation serves as the linchpin that ensures products reach their intended destinations efficiently, reflecting its pivotal role in modern commerce and customer satisfaction [4]. An innovative solution gaining attraction in the transportation sector is the utilization of smart contracts. Smart contract technology is a digital transaction protocol that seeks to establish agreements between multiple parties by automatically executing the contract's conditions. Smart contracts are disintermediated and typically transparent, offering reduced legal and transaction costs, enhanced business efficiency, and the possibility of anonymous transactions. Due to these attributes, smart contract technology is in high demand, particularly in the financial industry where it can mitigate the risks of fraud and nonpayment and enhance the quality of financial contracts with a certain degree of confidence, all without the involvement of intermediaries.

Smart contract technology has emerged as a robust tool for enhancing efficiency and security in various industries, including transportation [5]. However, its implementation in transportation faces specific challenges related to standardization, interoperability, reliable data inputs, and alignment with regulations and compliance requirements.

In general, smart contracts hold significant promise for building more reliable and effective transportation systems, ultimately improving product security, safety, and the timely delivery of shipments. This research paper investigates the potential applications of smart contracts in the field of transportation management. It explores a set of applications for smart contracts, including functions like shipment tracking, digital documentation management, and route optimization. Our analysis suggests that smart contracts have the capacity to revolutionize the transportation and logistics landscape, ushering in improvements characterized by swifter operations, cost-effectiveness, and heightened safety protocols.

This paper introduces a multi-layered smart contract framework tailored for logistics and smart transportation in Egyptian transportation, with the aim of addressing tracking and security issues inherent in transportation processes [6]. Smart contracts can enhance performance, reduce risks, mitigate fraud,

and improve efficiency, transparency, and tamper-proof capabilities [6]. These smart contracts are typically deployedon blockchain technology platforms like Ethereum, which provide a secure, transparent, and tamper-proof environment [8]. Blockchain's trusted and decentralized infrastructure is ideal for executing self-executing contracts. In the transportation industry, smart contracts can play a pivotal role in automating the execution of agreements among various parties involved, such as shippers and carriers [7]. Their applications encompass verifying the authenticity of goods, tracking shipments, ensuring compliance with rules and contractual obligations, automating processes, mitigating fraud risks, and enhancing security. Furthermore, they streamline payment procedures and automate financial transactions.

This paper is organized as follows: Section II provides a background on smart contracts. Section III introduces the related works in using smart contract in transportation. Section IV presents the proposed framework for Smart Contracts in transportation. Section V discusses the proposed framework. Implementation is given in Section VI. Results are analysed in Section VII. Finally, Section VIII presents conclusion.

II. BACKGROUND

Smart contracts are software components that automate the execution of contract terms and function as self-executing agreements. They address the limitations of traditional procedures and offer a performance level that reduces risks, minimizes fraud, and enhances efficiency, transparency, and security. Blockchain technology is an ideal platform for implementing smart contracts, providing a trusted and decentralized infrastructure for executing these agreements. Platforms such as Ethereum frequently utilize blockchain technology to offer a secure, transparent, and tamper-proof environment for smart contracts.

Smart contracts have diverse applications across industries, including finance, real estate, healthcare, transportation, and supply chain management. In the healthcare sector, they facilitate the secure exchange of patient data, while in real estate, they automate and secure property transfers. In supply chain management, they create transparent transaction records that assist managers in tracking products from their origin to their final delivery destination.

In transportation, smart contracts enable real-time tracking of goods, offering companies increased visibility and insights into their operations. These contracts streamline transportation processes by swiftly and securely automating various tasks, thereby enhancing overall efficiency. They facilitate agreements among parties such as shippers and carriers, verify the authenticity of goods, track shipments, and ensure compliance with rules and contractual obligations. Beyond process automation, smart contracts reduce fraud risks, enhance security, streamline payment procedures, and automate financial transactions, ultimately contributing to the development of more reliable and effective transportation systems. This improves product safety, ensures on-time delivery, and enhances overall security.

However, challenges exist in implementing smart contracts in transportation, including standardization and interoperability across different blockchain platforms and the need for reliable data inputs to ensure accurate contract execution. Additionally, smart contracts require careful preparation of contract terms and adherence to regulatory and compliance requirements.

A. Layers of Blockchain and Smart Contracts

The fundamental elements of blockchain and smart contracts can be categorized into five layers, as illustrated in Fig. 1: buyer and seller, applications, verification, blockchain types, and network.

First Layer: In this layer, Real Estate Registration and Documentation Authority employees are responsible for inputting buyer data.

Second Layer: In the second layer, developers create various applications for smart contracts, including back-end programs, front-end programs, and decentralized applications (DApps).

Third Layer: In this layer, individuals are responsible for transaction verification before registration, as well as the types of consensus algorithms employed.

Fourth Layer: In the fourth layer, we focus on the methods used for recording and storing transactions, including options like Private Blockchain, Public Blockchain, or Consortium Blockchain.

Fifth Layer: The final layer addresses the network used for registering smart contracts, which is a peer-to-peer network. This network consists of interconnected computer systems via the Internet, without a central server. It incorporates various cryptographic technologies, public-key encryption, and platforms for smart contracts such as NEAR, Ethereum, and Hyperledger. Transaction storage is managed through virtual storage for all participants on the platform.

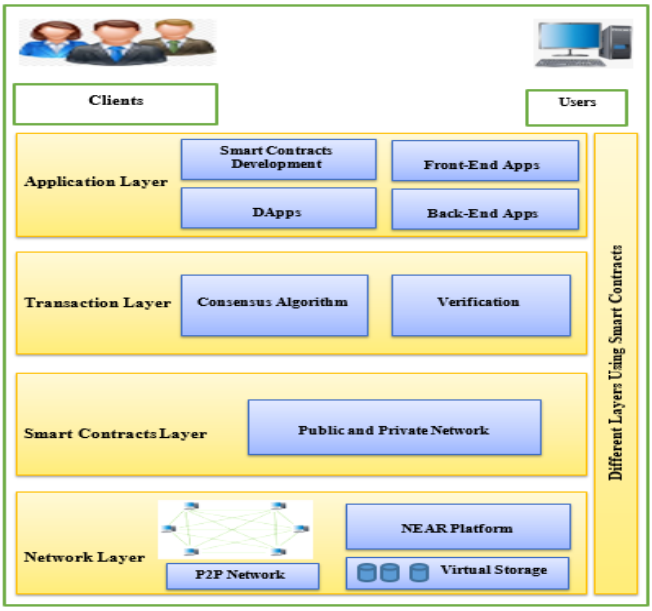


Fig. 1. Layers of smart contracts using blockchain.

B. Blockchain Types

1) Blockchain technology is categorized into three types: federated blockchain, public blockchain, and private blockchain. Each type is elaborated on as follows:

2) Public blockchain: A public blockchain allows anyone to join the network, verify transactions, and participate in consensus processes without disclosing their identities. Participants can send and receive transactions, validate blocks, and offer incentives to those who validate blocks. Public blockchains are unrestricted, and examples include platforms like Bitcoin and Ethereum [9].

3) Consortium or federated blockchain: Consortium or federated blockchains are not open to everyone; they are semi-private, with each participant having equal authority. Platforms like Corda and R3 are examples of this type [10].

4) Private blockchain: Private blockchains are networks accessible only to specific companies or banks, with stringent access control and a predefined set of nodes that are aware of each other. Examples include Monax and Multichain [11].

C. Blockchain Characteristics

Blockchain technology possesses a set of characteristics that make it a compelling solution for addressing various challenges [23].

1) Decentralized: One of blockchain's fundamental characteristics is its decentralization. It eliminates the reliance on centralized nodes to record and store data, opting for a decentralized, distributed ledger [12].

2) Immutable: Records stored in a blockchain are permanent and unalterable unless someone gains control of more than 51% of the network nodes simultaneously [24].

3) Open source: Most blockchain systems are open for public inspection, and users can harness blockchain technology to develop various applications, including smart contracts, cryptocurrencies, voting systems, healthcare solutions, insurance platforms, and supply chain tracking [13].

4) Transparent: Data records in a blockchain are transparent to all nodes, fostering trust during data updates.

5) Autonomous: Blockchain nodes can securely transfer data based on consensus algorithms [14].

6) Anonymity: Blockchain transactions provide anonymity by concealing the identity of the transaction executor and stored data [15].

7) Security: Blockchain offers enhanced security as there isn't a single point of breakdown capable of disrupting the entire network.

8) Increased capacity: The collaborative efforts of thousands of computers within a blockchain network enhance the overall network capacity [12].

D. Structure of Blockchain Data

The core data structure used in blockchain systems is a 'block,' with each block containing a distinct number known as a 'hash value.' This hash value is cryptographically generated using the SHA-256 algorithm and stored in the block's header. The cryptographic hash function takes any input string and produces a 64-bit hash value as output [15].

Key components of the block header include the block number and metadata, which encompass identification numbers, reference numbers, timestamps, and dates. Each block also references the previous block through its hash value or block number, forming a chain that allows all connected blocks to access the complete history [16]. The 'nonce' is the result of transaction processing with the assistance of a cryptographic hash algorithm (SHA-256) following encryption and data validation [25].

E. The Peer-to-Peer (P2P) Blockchain Network

The peer-to-peer network (see Fig. 2), often referred to as a P2P network, is a decentralized architecture in which computer systems are directly interconnected via the internet, eliminating the need for a central server. [21] This approach to networking follows a decentralized and distributed application model, dividing tasks or workloads among peers [26]. Participants in this network enjoy significant freedom in terms of connectivity. Additionally, the blockchain network relies on data cryptography technology for secure data transfer [22].

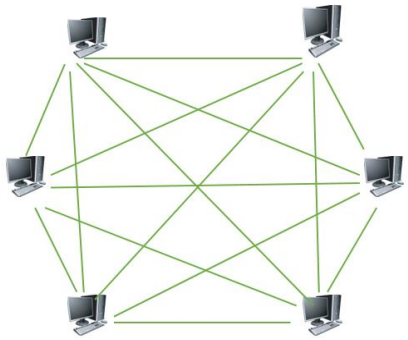


Fig. 2. Peer-to-peer blockchain network.

III. Related Works

In this section, the utility of smart contracts in transportation is investigated. The following subsections are categorized based on their applications of smart contracts to address transportation-related tasks. These tasks include intelligent electric vehicle systems, supply chain management, safety, security, traceability, mobility-as-a-service ecosystems, data management, real-time tracking, and cargo information [27].

A. Intelligent Electric Vehicle Systems and Road Safety

Several studies have demonstrated the effectiveness of using smart contracts in a blockchain-based electric vehicle transportation system. Byun et al. [29] developed an inference engine that enhances security and privacy by processing EV sensor data on the blockchain. This system efficiently manages the growing data flow from electric vehicles. On the other hand, Singh et al. [30] utilized smart contracts to establish trust in a blockchain-based Internet of Vehicles. Their approach addressed security and privacy concerns among untrusted vehicles, ultimately improving road safety and traffic efficiency. They also introduced measures to enhance scalability and transaction throughput. In addition to these works, Hou et al.

[42] explored the use of smart contracts within Decentralized Autonomous Organizations to streamline governance in Intelligent Transportation Systems (ITS). They proposed a DAO framework to simplify complex transportation systems. This innovation represents a promising future trend in ITS development.

However, these studies have focused on the security and trustiness of the blockchain based system. This gap in knowledge highlights the need for further research to explore the effectiveness of these systems to satisfy managerial needs of the stakeholders and the managers.

B. Supply Chain Management and Logistics

Several studies have demonstrated the effectiveness of using smart contracts in a blockchain-based supply chain management and logistics for automatic payment, cost reduction and transparency enhancement. For automatic payment in supply chains, Banerjee et al. [32] implemented smart contracts within an Intelligent Transport Management System based on blockchain (I-TMS) to automate toll-tax collection in supply chains. This approach ensures data security, transparency, and privacy through the decentralized and transparent nature of blockchain. Through the authentication of vehicle data, smart contracts enable automatic toll collection without the need for vehicle stops, resulting in fuel savings, improved time efficiency, and enhanced data security compared to traditional RFID-based systems. Furthermore, Baygin [33] employedsmart contracts in a blockchain-based fast shipping management architecture for local cargo networks to establish an automatic payment and approval mechanism. To reduce costs in transportation, smart contracts, as introduced by Philipp et al. [11], utilize blockchain technology to automate processes, reducing intermediaries and transaction costs within the transportation industry. [28] These contracts promote cross-organizational collaboration, particularly benefiting entrepreneurs and small and medium-sized enterprises by facilitating their entry into transnational supply chains. The abstract emphasizes the role of blockchain smart contracts in enhancing collaborative logistics and ensuring SME integration in sustainable maritime supply chains, highlighting their potential in transnational and multimodal supply chain environments.

In a study by Manimuthu et al. [17], smart contracts were employed in transportation to improve sustainable supply chain operations within the Autonomous Vehicles industry. This approach enhances traceability, transaction transparency, and sustainability, while also fostering the emergence of innovative AV business models. In personalized transportation, Chinaei et al. [38] utilized smart contracts within the transportation sector to promote the Mobility-as-a-Service concept. These contracts enable personalized transport service ownership through blockchain technology, providing a secure and private exchange of ownership. Their paper introduces a blockchain-driven MaaS platform that utilizes smart contracts to personalize service ownership, address congestion, and facilitate data trading among stakeholders, ultimately enhancing transportation service efficiency and convenience. To enhance transparency and security, Terzi et al. [63] employed smart contracts in supply chain management. These contracts enable product logging and tracing, ensuring unique item identification from factory to customer. Blockchain technology is also employed for user authentication and authorization, enhancing network security and trust while preserving data confidentiality. These smart contracts facilitate efficient interactions and trust amongsupply chain participants in real-life scenarios.

However, these studies have focused on partial aspects of the blockchain based systems. This gap in knowledge highlights the need for further research to explore the expandability and completeness of these systems to satisfy managerial needs of the stakeholders and the managers.

C. Safety, Security, and Traceability

Several studies have focused on ssafety, ssecurity and ttraceability of using smart contracts in a blockchain-based supply chain management. For example, in [18], Zichichi et al. employed smart contracts to coordinate data sharing in a smart transportation system using distributed ledgers such as Ethereum, IOTA, and IPFS. [19] These contracts enable secure and automated sensor data exchange while preserving privacy through Zero Knowledge Proof, fostering innovative smart services and societal benefits in user mobility. Similarly, Valencia-Payan et al. [36] utilized smart contracts to monitor and ensure coffee bean quality during transportation and storage via Blockchain technology. Their Hyperledger Fabric-based smart contract automates condition verification and enhances traceability, offering fast throughput and low latency suitable for real-world applications in the transportation and quality control of coffee beans. On the other hand, Imeri et al. [28] appliedsmart contracts to enhance security and traceability during the transportation of hazardous materials. Their conceptual approach employs blockchain-based smart contracts to securely share sensitive information while maintaining transparency and auditability. Authorized parties' access and interact with the data, ensuring the safe movement of dangerous goods. Moreover, Valchanov et al. [39] utilized smart contracts in transportation to improve safety in the transportation of special and hazardous cargo. Their model combines blockchain and IoT technology to record real-time sensor data and trigger smart contracts for stakeholder notifications when thresholds are breached. This expedites insurance claims and enhances safety. Finally, Mihelj et al. [41] developed a system for real-time identification of road traffic events and evaluation of source reputation using smart contracts in the transportation industry. This approach effectively detects manipulative behaviors, maintaining data integrity for safer transportation.

Generally, this research enhances trust and eliminate single authority control, ensuring reliable and secure traffic data collection and assessment.

However, these studies have focused only on safety, ssecurity, and ttraceability of the smart contracts in transportation systems. This gap in knowledge highlights the need for further research to explore the business requirements and the managerial needs of these systems.

IV. Mobility-as-a-Service (MaaS) Ecosystems

Several studies have focused on using smart contracts to serve as Mobility-as-a-Service (MaaS) ecosystems in transportation systems. For example, Chinaei et al. [37] employed smart contracts in transportation to advance the MaaS

concept using blockchain technology. These contracts enable personalized ownership of transport services, ensuring secure exchanges with privacy preservation. The paper introduces a blockchain-based MaaS platform, powered by smart contracts, to customize service ownership, manage congestion, and facilitate data trading among stakeholders, thus enhancing transportation efficiency and convenience. Additionally, Tallyn et al. [40] explored smart contract applications in transportation, with a specific focus on optimizing last-mile deliveries to improve precision, coordination, and accountability.

Furthermore, Karinsalo et al. [44] harnessed smart contracts in transportation to enable seamless and reliable transactions within Mobility-as-a-Service (MaaS) ecosystems, leveraging both blockchains and artificial intelligence. The proposed model, exemplified by Travel Token, employs blockchain-enabled smart contracts to store travel data, ensure dataintegrity, and automate value-sharing and compensation processes, thereby enhancing the efficiency and reliability of travel experiences while meeting industry demands.

However, these studies have focused on personalized ownership of transport services in transportation systems. This integration of smart contracts raises important questions about balancing automation with personal values, relationships, and individual agency within the transportation sector. This gap in knowledge highlights the need for further research to address evolving business needs in the MaaS field.

A. Data Management and Insurance

Several studies have focused on data management and insurance of smart contracts in transportation systems. Wanget al. [31] utilized smart contracts to establish an auditable payment and asset delivery protocol in transportation, ensuring transparent and secure transactions. This protocol enhances fairness, traceability, and security by leveraging blockchain technology, thereby addressing third-party payment and security issues. Balasubramaniam et al. [20] applied dimensionality reduction techniques and blockchain in transportation to present verifiable digital evidence while minimizing data for efficient storage on the blockchain. Smart contracts facilitate datasharing among insurance companies, enhancing data integrity within Intelligent Transportation Systems (ITS).

Furthermore, Quoc et al. [43] introduced the 'Safe Seller Safe Buyer' (SSSB) system in transportation, which eliminates intermediaries. Smart contracts include rules for conflict resolution and penalty enforcement, providing insurance-like protection. Lastly, Valencia-Payan et al. [28] applied smart contracts in goods transportation, with an emphasis on traceability and transparency in social selling.

However, these studies have focused on data management, insurance and safety in transportation systems without emphasizing on the global ecosystem interaction and collaboration. This gap in knowledge highlights the need for further research to address evolving business needs in the global ecosystem.

B. Real-Time Tracking and Decision-Making

Several studies have focused on role of smart contracts in transportation systems for real-time tracking and decision making. Arumugam et al. [34] applied smart contracts in transportation for advanced supply chain management and logistics, utilizing real-time IoT tracking and decision-making processes. These contracts allow users to establish trust and enforce conditions, ensuring visibility and traceability throughout the supply chain. Their proposed smart logistics system combines condition monitoring, logistics planning, and smart contracts to enhance accountability and efficiency in supply chain asset management. To enhance transportation management, Hasan et al. [35] employed smart contracts on the Ethereum blockchain for efficient supply chain management in transportation, with a focus on tracking and managing global trade shipments. These contracts govern stakeholder interactions, automate payments, and enforce predefined shipping conditions for IoT-equipped smart containers. This technology enhances transparency, security, and automation in complex logistics, with potential applications in vaccine supply chain management. Additionally, Arumugam et al. [42] used smart contracts in transportation to improve supply chain management and logistics through real-time tracking and decision-making with IoT devices. These contracts enable trust and enforceability, providing visibility and traceability throughout the supply chain. Their proposed smart logistics system integrates condition monitoring, logistics planning, and smart contracts to enhance accountability and efficiency in supply chain asset management. This streamlines document processing, reducing errors and manipulation, while enhancing efficiency and saving time and costs. Automation ensures the immutability, trust, and privacy of the blockchain, thereby improving the speed of the shipment process.

These studies have focused only on real-time tracking and decision making. This gap in knowledge highlights the need for further research to address other business needs in the transportation ecosystem.

C. Cargo Information and Traceability

Several studies have focused on role of smart contracts in transportation systems for tracking Cargo Information and Traceability. For example, Munari et al. [45] investigated smart contract usage in transportation, focusing on the Tradeless platform, a permissioned blockchain for real-time cargo data. Concerns arose regarding data transparency's anti-competitive effects and sensitive information exposure among competitors. The work explores blockchain and smart contracts' impact on transportation, highlighting changing transport patterns and reduced transmission of personal or sensitive data.

These studies have focused only on cargo information. This gap in knowledge highlights the need for further research to address other integrated business needs in the transportation ecosystem.

V. Proposed Framework for Smart Contracts Based Transportation

The transportation industry in the Arab World is one of the most lucrative sectors, with revenues increasing annually. Smart contracts have introduced new forms of interactions, suchas the reliable logging of transit traffic, which was previously unmanageable with traditional systems. The conventional transportation system in Egypt and other countries is slow and cumbersome, relying heavily on multiple intermediaries. By

integrating blockchain technology into the transportationsector, real-time tracking of transportation movements can be verified on a unified blockchain platform, ensuring transparency, high security, a decentralized environment, and a distributed ledger.

This paper proposes a formal method to ensure compliance with existing smart contract regulations and assign liability. The objective is to develop a distributed system that securely manages transactions to accelerate movement, streamline transfers, and address various transportation-related tasks. These tasks include supply chain management, safety, security, traceability, mobility-as-a-service ecosystems, data management, real-time tracking, and cargo information all without requiring vehicles to stop at toll plazas. This approach ensures data security, transparency, and privacy through the decentralized and transparent nature of blockchain. By authenticating vehicle data, smart contracts facilitate automatic toll collection without the need for stops, leading to fuel savings, improved time efficiency, and enhanced data security compared to traditional RFID-based systems.

Smart contracts, implemented on platforms such as NEAR and Ethereum, enable secure communication between parties and efficient tracking and tracing of RFID-tagged shipments using Internet of Things (IoT) and Ultra-High Frequency (UHF-RFID) sensors. This paper highlights an encouraging trend for the future development of Intelligent Transportation.

A. The Traditional Transportation System

In traditional transportation systems, customers typically visit the transport company in person to submit all necessary documents verifying their ownership of the goods. Afterward, the customer pays the required transportation fees and receives an invoice that includes the expected delivery date. One of the major drawbacks of traditional land transport is the lack of goods tracking, insufficient driver oversight, and inflexible scheduling. These issues can lead to damage to certain goods and increase the risk of theft due to the absence of tracking and a clear understanding of the different stages of the goods' journey, as illustrated in Fig. 3.

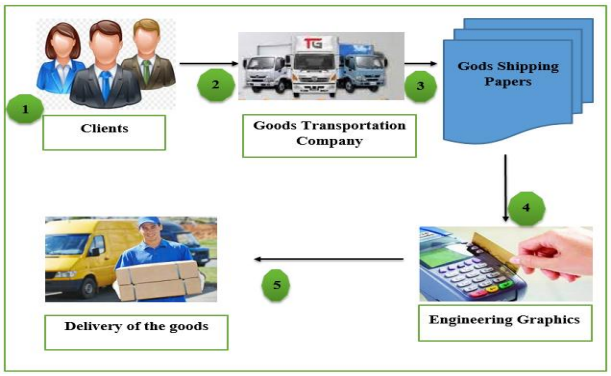


Fig. 3. Traditional transportation system in Egypt.

B. Smart Contract based Transportation System

This work proposes a framework for tracking transportation movements in Egypt using smart contracts and blockchain technology. The framework documents the transportation of goods through smart contracts facilitated by blockchain technology. Customers enter into an agreement with the transport company and provide all necessary documents verifying the ownership of the goods.

The shipping company then records the transaction on the smart contract, creating a new record and assigning a unique identifier to track the shipment. Blockchain technology simplifies shipping operations significantly by enabling direct and secure tracking of shipments at all stages. This is achieved through the integration of GPS (Global Positioning System, which provides positioning, navigation, and timing services) with sensors—electronic devices that monitor physical attributes such as temperature, pressure, distance, speed, torque, and acceleration.

The system consists of three chips: a space chip, a control chip, and a user chip, along with a mobile chip. These devices are attached to the goods cart, allowing precise monitoring of shipment details, tracking the status of shipments, and connecting the information to the smart contracts network. The proposed system aims to improve traceability, transparency, information security, and the immutability of data stored and exchanged during various operational processes in a decentralized manner, thereby eliminating the need for intermediaries or third parties, as illustrated in Fig. 4.

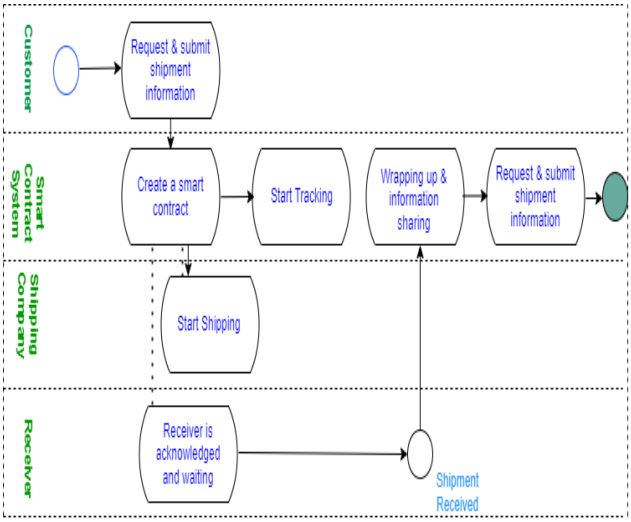


Fig. 4. The Proposed framework for tracking transportation in Egypt.

VI. Implementation

We will implement the proposed framework for smart contracts, designed to facilitate the registration and tracking of goods and commodities using blockchain technology. The protocol has been developed and executed on the NEAR platform. Smart contracts are integral in establishing a distributed, secure, and decentralized ledger to safely record all assets and transactions between parties. Our model is developed using Assembly Script and Rust programming languages and operates on the NEAR platform, a decentralized application based on blockchain technology, where it meticulously stores all transaction records related to the goods tracking process.

During the design phase, the necessary input forms are defined and translated into code programs. These screens include various functions such as user login, registration of buyer and seller information, payment processing, document

review and validation, smart contract creation, TXN hash, and ID property generation. The NEAR platform is notable for offering the lowest transaction fees, approximately $0.0001 per transaction, and supports up to 100,000 transactions per second (TPS), outperforming many other platforms.

A. Testing and Assessment

This study addresses two types of assessments: statistical analysis and performance evaluation.

B. Statistical Analysis

The primary objective of the surveys is to examine the feasibility of adopting smart contracts based on blockchain technology as an alternative to the conventional shipment registration system in Egypt. This research employs a questionnaire study to validate the proposed model's effectiveness and a survey to assess the impact of various factors on the adoption of blockchain-integrated smart contracts within the Egyptian shipment registration system.

The questionnaire evaluates the influence of independent variables, such as data availability, location accessibility, ease of transaction verification, transaction comprehension, data accuracy, data immutability, cost reduction, affordable contract execution, and improvements in process efficiency, on the dependent variable: the integration of blockchain technology with the Egyptian cargo tracking infrastructure.

The research sample was carefully selected based on relevant criteria and included individuals with a keen interest in blockchain and smart contracts, IT specialists, and professionals involved in the shipping industry. The proposed framework was introduced and practically demonstrated to each member of the research sample, followed by initial testing to ensure the model's practical applicability.

The sample consisted of 127 participants interested in the study's subject matter. The study on hand categorized the sample into groups of questions concerning blockchain, smart contracts, the questionnaire format, and the total number of questions (15). This section presents the conclusions drawn from the survey data analysis, which was conducted using the Statistical Package Table II summarizes that.

Table I illustrates a Comparison of Some of the most popular Smart contract platforms.

This section discusses a comparison of the five most popular smart contract platforms Hyperledger Fabric, Ethereum, NEAR, Solana, and Cardano, highlighting their main characteristics like Transaction per second (TPS), Average transaction fee, Network type, Programming Language, Market Capital, and highlighting the importance of using the NEAR platform Table I summarizes that.

TABLE I. PLATFORM COMPARISON

Description	Ethereum	NEAR	Hyperledger Fabric	Solana	Cardano
Used Language	Solidity	Assembly Script, Rust	JavaScript, Go	C, Rust	Plutus
Foundation	Ethereum developers	NEAR foundation	Linux foundation	Solana foundation	Cardano foundation
Ledger type	Permissionless	Permissioned	Permissioned	Permissionless	Public
Cryptocurrency	ETH	NEAR	No Token	SOL	ADA
Coin market cap 5/2023	$ 1,08	$ 2.21	Not Applicable	$ 35.09	$ 0.48
Consensus algorithm	Proof of work (PoW)	proof of stake (PoS)	Crash Fault Tolerance (CFT)	Proof of stake (PoS-PoH)	Proof of stake(PoS)
Open Source	Yes	Yes	Yes	Yes	Yes
Using	Easy to use	Easy to use	Easy to use	Easy to use	Easy to use
Transactions per second (TPS)	27 / TPS	100,000 / TPS	3,500 / TPS	3,240 / TPS	250 / TPS
Average transaction fees	$ 0.7962	$ 0.0001	No transaction fees	$ 0.00025	$ 0.4
Network Type	Public Network	Public and Private Network	Private Network	Public Network	Public Network
Market Capital	$ 225,052,960	$ 1,756,564	$ 1,148,790	$ 8,532,436	$ 13,652,673
Date released	2013	2020	2016	2017	2013

Sciences (SPSS) as the analytical tool.

TABLE II. SUMMARY OF DEMOGRAPHIC PROFILE OF RESPONDENTS

Variable	Category	Frequency	Percentage (%)
Gender	Male	84	66%
	Female	43	34%
	Total	127	100%
Age	51 to 60	21	17%
	41 to 50	42	33%
	31 to 40	27	21%
	20 to 30	37	29%
	Total	127	100%
Specialty	Information Technology Specialists	37	29%
	Interested in buying and selling	29	23%
	Employees	31	24%
	Others	30	24%
	Total	127	100%

TABLE III. FREQUENCIES

Q	Strongly Agree (100%)		Agree (75%)		Strongly Agree	Neutral	Neither Agree nor Disagree (50%)		Disagree (25%)		Strongly Disagree (0%)	
	Frequency	Row N%	Frequency	Row N%			Frequency	Row N%	Frequency	Row N%	Frequency	Row N%
Q1	70	55%	50	39%	120	94%	94%	6%	7	6%	0	0%
Q2	68	54%	45	35%	113	89%	89%	11%	13	10%	0	0%
Q3	70	55%	44	35%	114	90%	90%	10%	13	10%	0	0%
Q4	50	39%	57	45%	107	84%	84%	16%	19	15%	1	1%
Q5	74	58%	48	38%	122	96%	96%	4%	5	4%	0	0%
Q6	35	28%	65	51%	100	79%	79%	21%	25	20%	1	1%
Q7	70	55%	47	37%	117	92%	92%	8%	8	6%	2	2%
Q8	51	40%	62	49%	113	89%	89%	11%	13	10%	1	1%
Q9	51	40%	67	53%	118	93%	93%	7%	7	6%	2	2%
Q10	11	9%	33	26%	44	35%	35%	65%	24	19%	35	28%
Q11	9	7%	29	23%	38	30%	30%	70%	46	36%	19	15%
Q12	29	23%	43	34%	72	57%	57%	43%	46	36%	7	6%
Q13	37	29%	15	12%	52	41%	41%	59%	60	47%	9	7%
Q14	7	6%	14	11%	21	17%	17%	83%	39	31%	37	29%
Q15	52	41%	60	47%	112	88%	88%	12%	9	7%	1	1%

VII. RESULTS AND FINDINGS

The survey findings reveal that a significant majority of respondents hold positive views regarding the adoption of smart contracts. Notably, 96% of participants found it convenient to access data and monitor shipments and logistics using smart contracts and cryptography. Additionally, 92% acknowledged that smart contracts enhance security and privacy for all stakeholders involved. A substantial 84% agreed that the integration of smart contracts eliminates intermediaries acting as third parties between sellers and buyers. Furthermore, 93% concurred that using Transportation in blockchain technology in conjunction with Supply chains and smart contracts helps mitigate tax evasion, as illustrated in Fig. 5.

The survey results indicate that 90% of respondents believe that smart contracts effectively regulate the shipping tracking process, while 89% agree that employing blockchain technology alongside smart contracts helps reduce tax evasion. Additionally, 79% observed significant changes in the electronic shipping registration process compared to traditional methods. Notably, 17% of respondents acknowledged the inefficiency and sluggishness of traditional contract registration procedures. The questionnaire results strongly support the notion that incorporating blockchain technology and smart contracts into the shipping registration process in Egypt would provide various benefits, including rapid access to transaction information, streamlined verification processes, improved data accuracy, increased reliability, and significant cost reduction.

The survey findings suggest that adopting blockchain technology in conjunction with smart contracts for shipping registration in Egypt, offers advantages such as fast access to shipment location data, simplified transaction verification, improved data precision, enhanced reliability, and substantial cost savings.

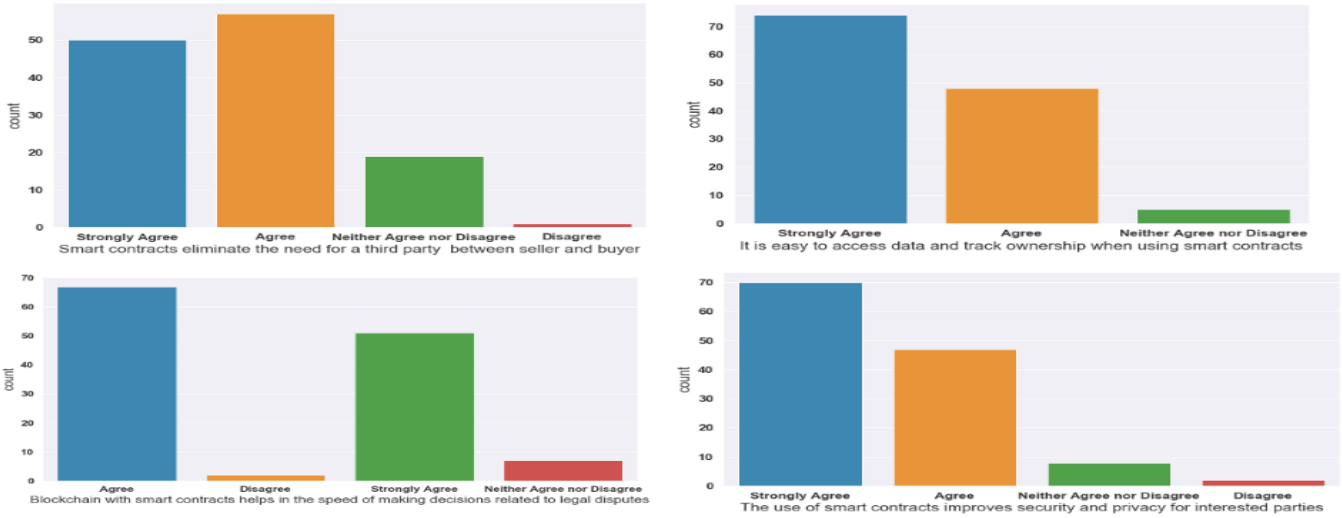


Fig. 5. Analysis results.

To further evaluate the model's performance, the proposed smart contract model was deployed on the NEAR platform, and its performance was compared to other smart contract platforms. The smart contract was written in Assembly Script on the NEAR Web3 platform. The model effectively addresses shipping registration challenges in Egypt and demonstrates superior attributes compared to Ethereum-based smart contracts. It successfully resolves issues related to registration time, sequencing characteristics, cost-efficiency, and transaction speed. While smart contract development can be performed on various blockchain platforms, the choice of platform is influenced by decentralized application storage techniques.

In our view, the positive response to smart contracts demonstrates the significant potential of this approach to transform Egypt's transportation sector. The high percentages in the results confirm the practicality of integrating smart contracts. Additionally, the reduction in tax evasion and the increased registration efficiency offer promising implications for economic governance and operational sustainability. However, challenges related to classical bureaucratic processes still need to be addressed Table III summarizes that.

VIII. Conclusion

This study demonstrates how smart contracts and blockchain technology can significantly improve transportation and logistics in Egypt. It addresses key challenges such as congestion, inefficiency, and a lack of transparency in conventional systems. Our smart contract framework, powered by blockchain technology, offers a secure and decentralized method for transportation management, streamlining processes and addressing logistical tasks such as supply chain management and real-time tracking. The framework, implemented on the NEAR platform, outperforms traditional methods by delivering improved information accessibility and enhanced privacy. The features of the framework include the integration of technologies, user-focused design, decentralization, and the scalability of the NEAR model. Survey results indicate a 96% increase in convenience for data access, a 92% improvement in privacy and security, and an 84% reduction in the need for intermediaries. Other results indicate a 93% increase in the mitigation of tax evasion and a 93% improvement in shipping registration. While the framework provides promising results, further research is needed to improve scalability and address potential regulatory challenges. Future work could recommend integrating analytics and machine learning to further optimize logistics or applying the model to other industries. In addition, further investigation can be conducted into IoT integration, regulatory frameworks, cross-border applications, environmental impact studies, blockchain scalability, and cybersecurity risk evaluation. In conclusion, smart contracts and blockchain technology hold significant potential for transforming transportation management, particularly in Egypt."

Acknowledgment

This work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under grant number (UJ-23-DR-150). Therefore, the authors thanks the University of Jeddah for its technical and financial support.

References

[1] Chen, X. and X. Zhang, Secure electricity trading and incentive contract model for electric vehicle based on energy blockchain. IEEE access, 2019. 7: p. 178763-178778.

[2] Tijan, E., et al., Blockchain technology implementation in logistics. Sustainability, 2019. 11(4): p. 1185.

[3] Chang, S.E., Y.-C. Chen, and M.-F. Lu, Supply chain re-engineering using blockchain technology: A case of smart contract based tracking process. Technological Forecasting and Social Change, 2019. 144: p. 1-11.

[4] Nanda, S.K., S.K. Panda, and M. Dash, Medical supply chain integrated with blockchain and IoT to track the logistics of medical products. Multimedia Tools and Applications, 2023: p. 1-23.

[5] Humayun, M., et al., Emerging smart logistics and transportation using IoT and blockchain. IEEE Internet of Things Magazine, 2020. 3(2): p. 58-62.

[6] Liao, D.-Y. and X. Wang. Applications of blockchain technology to logistics management in integrated casinos and entertainment. in Informatics. 2018. MDPI.

[7] Li, M., L. Shen, and G.Q. Huang, Blockchain-enabled workflow operating system for logistics resources sharing in E-commerce logistics real estate service. Computers & Industrial Engineering, 2019. 135: p. 950-969.

[8] Rohr, J.G., Smart contracts and traditional contract law, or: the law of the vending machine. Clev. St. L. Rev., 2019. 67: p. 71.

[9] Guegan, D., Public blockchain versus private blockhain. 2017.

[10] Yang, R., et al., Public and private blockchain in construction business process and information integration. Automation in construction, 2020. 118: p. 103276.

[11] Patel, V., et al., A review on blockchain technology: Components, issues and challenges, in ICDSMLA 2019. 2020, Springer, Singapore. p. 1257-1262.

[12] Atlam, H.F., et al., Blockchain with internet of things: Benefits, challenges, and future directions. International Journal of Intelligent Systems and Applications, 2018. 10(6): p. 40-48.

[13] EDU, W.S.O., PropTech 3.0: the future of real estate. Retrieved May, 2021. 10: p. 2018-07.

[14] Fan, L., F. Cronemberger, and J.R. Gil-Garcia, Using Blockchain Technology to Manage IoT Data for Smart City Initiatives: A Conceptual Framework and Initial Experiments Based on Smart Contracts, in Beyond Smart and Connected Governments. 2020, Springer. p. 85-108.

[15] Lin, W., et al., Blockchain technology in current agricultural systems: from techniques to applications. IEEE Access, 2020. 8: p. 143920-143937.

[16] Kan, J., S. Chen, and X. Huang. Improve blockchain performance using graph data structure and parallel mining. in 2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN). 2018. IEEE.

[17] Manu, M., et al., Blockchain Components and Concept, in Blockchain Technology and Applications. 2020, Auerbach Publications. p. 21-50.

[18] Vacca, A., et al., A systematic literature review of blockchain and smart contract development: Techniques, tools, and open challenges. Journal of Systems and Software, 2020: p. 110891.

[19] Terzi, S., Zacharaki, A., Nizamis, A., Votis, K., Ioannidis, D., Tzovaras, D., & Stamelos, I. (2019). Transforming the Supply-Chain Management and Industry Logistics with Blockchain Smart Contracts. In Proceedings of the 23rd Pan-Hellenic Conference on Informatics (pp. 9-14). Association for Computing Machinery. DOI: 10.1145/3368640.3368655.

[20] Xu, X., I. Weber, and M. Staples, Architecture for blockchain applications. 2019: Springer.

[21] Mehar, S., Zeadally, S., Remy, G., & Senouci, S. M. (2014). Sustainable transportation management system for a fleet of electric vehicles. IEEE transactions on intelligent transportation systems, 16(3), 1401-1414.

[22] Bauwens, M., V. Kostakis, and A. Pazaitis, Peer to peer. 2019: University of Westminster Press.

[23] Stefansson, G., & Lumsden, K. (2009). Performance issues of smart transportation management systems. International Journal of productivity and performance management, 58(1), 55-70.

[24] Tarapiah, S., Atalla, S., & AbuHania, R. (2013). Smart on-board transportation management system using gps/gsm/gprs technologies to reduce traffic violation in developing countries. International Journal of Digital Information and Wireless Communications (IJDIWC), 3(4), 96-105.

[25] Zhu, F., Li, Z., Chen, S., and Xiong, G. (2016). Parallel transportation management and control system and its applications in building smart cities. IEEE Transactions on Intelligent Transportation Systems, 17(6), 1576-1585.

[26] Adler, J. L., and Blue, V. J. (2002). A cooperative multi-agent transportation management and route guidance system. Transportation Research Part C: Emerging Technologies, 10(5-6), 433-454.

[27] Karbassi, A., and Barth, M. (2003, June). Vehicle route prediction and time of arrival estimation techniques for improved transportation system management. In IEEE IV2003 intelligent vehicles symposium. Proceedings (Cat. No. 03TH8683) (pp. 511-516). IEEE.

[28] Ferguson, E. (1990). Transportation demand management planning, development, and implementation. Journal of the American Planning Association, 56(4), 442-456.

[29] Masek, P., Masek, J., Frantik, P., Fujdiak, R., Ometov, A., Hosek, J., and Misurec, J. (2016). A harmonized perspective on transportation management in smart cities: The novel IoT-driven environment for road traffic modeling. Sensors, 16(11), 1872.

[30] P.W; Byun, Y.-C. Smart Contract Centric Inference Engine For Intelligent Electric Vehicle Transportation System. Sensors 2020, 20, 4252. https://doi.org/10.3390/s20154252

[31] P. K. Singh, R. Singh, S. K. Nandi, K. Z. Ghafoor, D. B. Rawat and S. Nandi, "Blockchain-Based Adaptive Trust Management in Internet of Vehicles Using Smart Contract," in IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 6, pp. 3616-3630, June 2021, doi: 10.1109/TITS.2020.3004041.

[32] Das, D., Banerjee, S., and Chatterjee, P. (2022). Design and development of an intelligent transportation management system using blockchain and smart contracts. Cluster Computer 25, 1899–1913 (2022). https://doi.org/10.1007/s10586-022-03536-z

[33] Mehmet Baygin, Orhan Yaman, Nursena Baygin, and Mehmet Karakose, "A blockchain-based approach to smart cargo transportation using UHF RFID," Expert Systems with Applications, Volume 188, 2022, 116030, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2021.116030

[34] S. Wang, X. Tang, Y. Zhang and J. Chen, "Auditable Protocols for Fair Payment and Physical Asset Delivery Based on Smart Contracts," in IEEE Access, vol. 7, pp. 109439-109453, 2019, doi: 10.1109/ACCESS.2019.2933860.

[35] Haya Hasan, Esra AlHadhrami, Alia AlDhaheri, Khaled Salah, and Raja Jayaraman, "Smart contract-based approach for efficient shipment management," Computers & Industrial Engineering, Volume 136, 2019, pp. 149-159. ISSN 0360-8352.

[36] C. Valencia-Payan, J. F. Grass-Ramírez, G. Ramirez-Gonzalez and J. C. Corrales, "A Smart Contract for Coffee Transport and Storage With Data Validation," in IEEE Access, vol. 10, pp. 37857-37869, 2022, doi: 10.1109/ACCESS.2022.3165087.

[37] A. Imeri and D. Khadraoui, "The Security and Traceability of Shared Information in the Process of Transportation of Dangerous Goods," 2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS), Paris, France, 2018, pp. 1-5, doi: 10.1109/NTMS.2018.8328751.

[38] Chinaei, Mohammad Hossein, Taha Hossein Rashidi, and Travis Waller. "Digitally transferable ownership of mobility-as-a-service systems using blockchain and smart contracts." Transportation Letters, vol. 15, no. 1, 2023, pp. 54-61. ISSN 1942-7867.

[39] Volchenkov, H., & Aleksieva, V. (2022). Blockchain and IoT integration for smart transportation. Journal of Physics: Conference Series, 2339(1), 012012.

[40] Tallyn, E., Revans, J., Morgan, E., Fisken, K., and Murray-Rust, D. (2021). Enacting the Last Mile: Experiences of Smart Contracts in Courier Deliveries. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21) (pp. 639). Association for Computing Machinery.

[41] Mihelj J, Zhang Y, Kos A, and Sedlar U. Crowdsourced Traffic Event Detection and Source Reputation Assessment Using Smart Contracts. Sensors. 2019; 19(15):3267. https://doi.org/10.3390/s19153267

[42] S. S. Arumugam et al., "IOT Enabled Smart Logistics Using Smart Contracts," 2018 8th International Conference on Logistics, Informatics and Service Sciences (LISS), Toronto, ON, Canada, 2018, pp. 1-6, doi: 10.1109/LISS.2018.8593220.

[43] Quoc, K.L. et al. (2022). SSSB: An Approach to Insurance for Cross-Border Exchange by Using Smart Contracts. In: Awan, I., Younas, M., Poniszewska-Marańda, A. (eds) Mobile Web and Intelligent Information Systems. MobiWIS 2022. Lecture Notes in Computer Science, vol 13475. Springer, Cham. https://doi.org/10.1007/978-3-031-14391-5_14.

[44] A. Karinsalo and K. Halunen, "Smart Contracts for a Mobility-as-a-Service Ecosystem," 2018 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C), Lisbon, Portugal, 2018, pp. 135-138, doi: 10.1109/QRS-C.2018.00036.

[45] Valencia-Payan, C., Grass-Ramírez, J. F., Ramirez-Gonzalez, G., & Corrales, J. C. (2023). Smart Contract to Traceability of Food Social Selling. Computers, Materials & Continua, 10.32604/cmc.2023.031554.

AUTHORS' PROFILE

Dr. Abdullah Alshahrani: Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Jeddah 23218, Saudi Arabia. Dr. Alshahrani is an Associate Professor, Department of Computer Science and Artificial Intelligence, University of Jeddah. Holds a PhD from the Department of Electrical Engineering and Computer Science at the Catholic University of America in Washington DC, USA with distinction with honors in a delicate specialization in conserving and prolonging the life of the sensor energy in wireless networks using several theories in the field of artificial intelligence and medical genetic theory in 2018. Graduated from the Australian La Trobe University with a master's degree in computer science in (2010) during the academic journey to obtain a master's degree, I was a member of Australian Computer Science committee. I attended many specialized courses in management and courses in team building, time management and project management. Published scientific papers in many prestigious journals in various disciplines of computer engineering, Networking and artificial intelligence.

Prof. Ayman E. Khedr: Professor Khedr is a Professor of Information Systems and currently a professor at the University of Jeddah. I have been the vice dean of post-graduation and research and the head of the Information Systems Department in the Faculty of Computers and Information Technology, at Future University in Egypt. I am a professor in the Faculty of Computers and Information, at Helwan University in Egypt. I have previously worked as the general manager of the Helwan E-Learning Centre. My research is focused on the themes (scientific) data and model management, Data Science, Big Data, IoT, E-learning, Data Mining, and Cloud Computing.

Prof. Mohamed A. Belal: Professor Belal is a Professor of Computer Science, specializing in Computational Intelligence, at the Faculty of Computers and Artificial Intelligence, Helwan University, Egypt. He is a consultant to the Ministry of Communications and Information Technology for AI-Based Semantic Search. His administrative roles have included serving as the Dean of the Faculty of Computers and Artificial Intelligence at Helwan University from 2012 to 2015 and at Beni-Suef University from 2016 to 2020. He was the Vice President of Beni-Suef Technological University from 2020 to 2021 and is currently the Vice President of the 6th. October Technological University, Egypt.

Dr. Mohamed Saleh Darwish: Dr. Saleh holds his PhD from the Department of Business Information Systems, Faculty of Commerce and Business Administration, Helwan University, Cairo, Egypt. Currently, He is working as the Executive Director at North Africa for Real Estate Investment. Lives in Cairo, Egypt.

Enhancing Educational Outcomes Through AI Powered Learning Strategy Recommendation System

Daminda Herath, Chanuka Dinuwan, Charith Ihalagedara, Thanuja Ambegoda
Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka

***Abstract*—In order to develop intelligent learning recommendation systems, the work identifies the employment of artificial intelligence (AI) techniques, particularly in the educational data mining (EDM) field. The aggregation of such educational data into an efficient analytical system could also assist as an interesting means of education for the students. In fact, it could ultimately advance the direction of education. Sophisticated machine learning methods were employed to analyze various data types, including educational, socioeconomic, and demographic data, to predict student success. In this research, Logistic Regression (LR), Random Forest (RF), Support Vector Machines (SVM), CatBoost, and XGBoost algorithms were considered to build prediction models using a dataset encompassing a wide range of student traits. Robust evaluation metrics, including precision, recall, accuracy, and F1-score, were used to gauge model effectiveness. The results highlighted that RF was the best with accuracy, precision, and recall. Then, a rule engine was built to enhance the system by finding the most efficient learning tactics for students based on their expected future performance. The proposed AI-based personalized recommendation tool shows a substantial step towards enhancing educational decisions. This solution facilitates educators in creating student academic assistance interventions by offering individualized, data-driven learning strategies.**

Keywords*—Artificial intelligence; educational data mining; educational strategies; machine learning; personalized recommendation; student performance prediction

I. Introduction

The incorporation of online learning environments and traditional classroom settings within educational institutions is progressively supported by an increasing volume of educational data. These data-rich settings offer substantial opportunities but also pose considerable challenges as institutions struggle to effectively bind them to enhance educational outcomes [1]. A key challenge is the collection and analysis of massive datasets to produce meaningful insights that can significantly influence educational strategies [2].

The science of discovering novel patterns in vast volumes of data is known as data mining. Moreover, finding knowledge in educational data that could significantly influence research methods in the field of education is the goal of educational data mining, or EDM [3]. In the past few years, researchers in this field have been eager to get valuable and practical insights, especially about student performance [4]. The aim to assist educational establishments in making choices and improving their teaching strategies is the motivational drive. In addition to help students overcome obstacles to learn and enhance their academic performance [5].

Failure and dropout rates are still rising despite technical advancements in e-learning that have undoubtedly had an impact on student performance [6]. This situation demands for a complete examination of a variety of influencing factors, including demographics, personal experiences, academic data, institutional surroundings, socioeconomic standing, and individual self-efficacy [7]. In fact, understanding the relationship between these factors is critical for evolving effective learning strategies and supporting academic accomplishment [8].

Despite the availability of extensive educational data, institutions face challenges in effectively utilizing this information to predict student performance and develop personalized learning strategies. Current systems frequently do not account for the diverse socioeconomic, demographic, and academic factors that affect student success. Consequently, dropout rates remain high, and students do not receive the individualized support necessary to enhance their academic outcomes.

In this study, the research objectives are; (1) to identify the key demographic, socioeconomic, and academic factors that influence student performance in higher education; (2) To apply advanced data mining techniques and machine learning models to predict student outcomes, specifically focusing on graduation, dropout, and enrollment statuses; and (3) To design and implement an AI-powered learning strategy recommendation that delivers personalized educational interventions based on predicted student performance.

The main contributions and implications of this study include the identification of key factors influencing student performance, the development of a predictive model utilizing machine learning algorithms to forecast student outcomes, and the introduction of an AI-powered personalized recommendation system that suggests individualized learning strategies based on performance predictions. This system enables educators to provide tailored interventions, which may help reduce dropout rates and enhance academic outcomes, while also delivering actionable insights for institutional policies and resource allocation. Furthermore, the system serves as a practical tool for educators to intervene early in students' academic journeys, allowing students to benefit from personalized learning strategies.

All in all, the structure of this paper is as follows: literature review in Section II, research methodology in Section III, results and evaluation in Section IV, discussion in Section V and finally conclusion in Section VI.

II. Related Works

Numerous research studies are included with data mining techniques with classification and clustering [9], prediction modeling [10], and AI-driven learning [13]. It is therefore anticipated that by using these techniques on educational data, useful knowledge and information will be extracted, raising the standards of the educational system. It is important to identify learning strategies and interventions that have been proved beneficial for students to give recommendations [14].

In previous studies, it was common to launch classification and clustering approaches working with educational data. In the study of these techniques, Song et al. [9] used the support vector machine and k-means clustering, to group students based on their performance and label them as successful or not. The student record embodied 41 variables, the components of which were delinquency behaviors, academic subjects, and demographics. SVM is a process of predicting results according to clusters, while being a clustering algorithm, k-means is used to set students' socioeconomic classes and effective learning into categories.

On the other hand, Alhassan et al. [10] applied classification algorithms to their research study to explore the affinity of grades obtained in coursework and online active data in student performance. As per the research, assessment grades are the most valuable factors that can be used to predict academic performance. Besides, modeling that uses assessment grades and data on online activity also outperform others. In fact, the two algorithms that worked best were Random Forest and Decision Tree. The study emphasizes the important of incorporating instructional technologies into learning environments.

Additionally, Mohamed Nafuri et al. [11] aimed to develop clustering analysis to classify the academic performance of students in the context of Malaysian higher education and develop graduation rates. The dataset, for instance, entailed data on gender, education levels, vocational exposure, co-curricular activities, and awards. Further, techniques such as k-means, BIRCH, and DBSCAN were used when data had been prepared. Also, the best model was KMoB, which was the optimized k-mean model, and with that, five clusters of student performance were identified. Despite the fact that such stigmatization can affect the academic performance as well as the social status of individual group members, their insights showed the necessity of developing individualized educational plans with data-driven approach and predictive analytics.

In addition to the above, Queiroga et al. [12] relied on data mining techniques and evaluated data obtained from 4529 undergraduate students. The system established it by utilizing the Virtual Learning Environment, survey, and academic system data to forecast student achievement. The research revealed as highly predictive in terms of amenability to treatment, which had caused certain elements to emerge as seemingly playing a determining role, including education, subject enrollment, and neighborhood. The results were followed by specific policies of the institutions, including recruitment and reporting of resources in addition to student tracking. Besides, they also used Python with NumPy, Pandas, and Scikit-learn libraries to implement a model. This work disclosed the potential of AI to personalized learning experiences based on individual student data.

Moreover, using data mining techniques like, Naïve Bayesian, Artificial Neural Network, Support Vector Machine and Decision Tree Classifier, Li et al. [13] inspected the contributing factors to students' performance. The characteristics like attendance, gender, and nationality were detected with the help of the information of the Kalboard360 E-learning platform. The results showed that support vector machines algorithm which takes the behavioral features such as resource use and participation into consideration as the most powerful at predicting performance. The researchers explored how to integrate behavioral data in such a way as to increase accuracy of forecasting and make learning more relevant for an individual learner.

Furthermore, Ouyang et al. [14] considered the possibility to describe how AI is being used in education to navigate educational environments and to predict academic success, so that educational failure is less, and the learning process is more effective. This study, which was conducted across an online engineering course, discovered that the adoption of AI learning analytics has concrete effects like collaboration, students' satisfaction, and engagement. In fact, this paper presented the research cases that illustrated the application of AI to educational reforms, in online learning, and higher education.

Moreover, for an efficacious academic achievement and to build personal effectiveness of academics in order to possibly minimize the likelihood of academically inadequate students. Renzulli [15] has suggested several effective strategies for learning; active involvement into study content, self-testing, note-taking, time management, flexed learning, distributed practice, interleaved practice, seeking somebody, self-regulation, and using university tools and resources. This paper reveals the potential of AI technology to execute authentic applications which could assist teachers to perform well academically.

In summary, all these studies show that data mining techniques now play a crucial role in educational research through providing valuable feedback on educational data and this results in the quality improvement of the educational system. Even though such attempts have been a few, gaining access to "real-time" data and recommending learning strategies based on student performance are not yet addressed sufficiently.

III. Research Methodology

Upstream industry pioneers have produced a standard called the Cross Industry Standard Process for Data Mining (CRISP-DM). It is equally and widely utilized in a variety of domains [16]. The six stages of this standard are; Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment are its six stages [17]. In fact, its broad and thorough application in the sphere of education is particularly noteworthy because it extends beyond other domains of use [18].

In this study, a predictive model creation strategy was developed using the CRISP-DM model, providing a

comprehensive structure for project implementation. Fig. 1 shows the complete CRISP-DM steps.

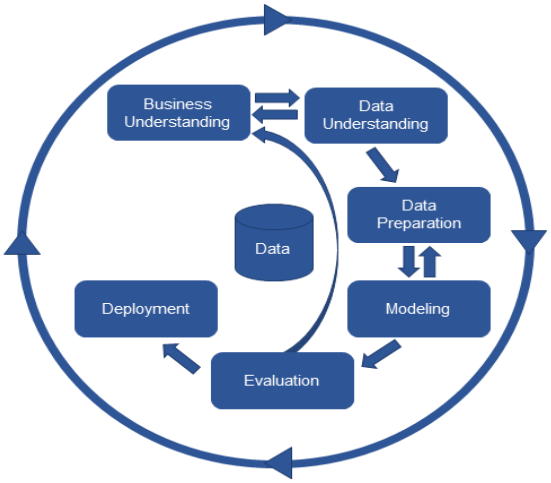


Fig. 1. Complete CRISP-DM step [17].

A. Business Understanding

This paper suggested a predictive model and recommendation system for student academic performance using educational statistics, aiming to enhance communication between educators and students, potentially enhancing learning environments [19].

As part of the methodology, attributes anticipated to have a direct impact on students' status were selected. To achieve this, a thorough investigation was conducted, followed by the construction of a figure (as shown in Fig. 2) demonstrating the relationship between these attributes and the students' attained status. Students were then classified into three categories (graduate, dropout, and enrolled) using various supervised learning techniques. Ultimately, learning tactics were suggested by the recommendation system based on the students' standing.

B. Data Understanding

The Kaggle project "Predict students' dropout and academic success" [20] provided the dataset used in this investigation. This dataset offers a comprehensive picture of the students enrolled in different undergraduate programs at a university. It contains information on social-economic characteristics, academic achievement, and demographics that can be utilized to examine potential determinants of academic success and student dropout. The dataset comprises 4424 instances, each associated with 35 attributes. The target attributes were graduate, dropout, and enrolled. The Table I shows description of the dataset including features and their type.

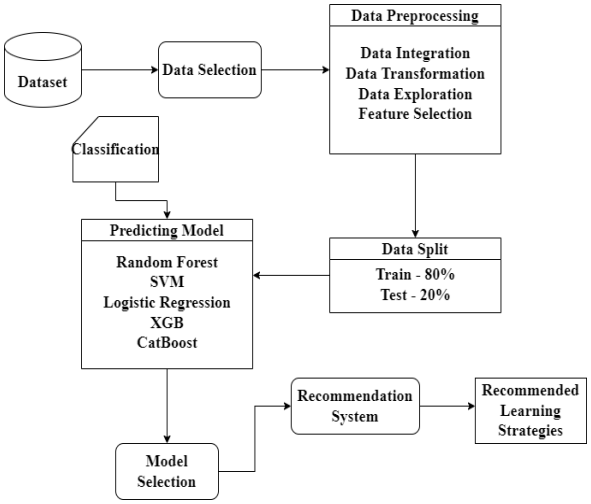


Fig. 2. Research methods for the prediction model.

TABLE I. DESCRIPTION OF THE DATASET INCLUDING FEATURES AND THEIR TYPE

Category	Feature	Description	Type
Demographic Data	Marital status	The marital status of the student	Categorical
	Gender	The gender of the student	Categorical
	Nationality	The nationality of the student	Categorical
	Age at enrollment	The age of the student at the time of enrollment.	Numerical
	International	Whether the student is an international student	Categorical
Academic Data	Course	The course taken by the student	Categorical
	Daytime/evening attendance	Whether the student attends classes during the day or in the evening	Categorical
	Curricular unit's 1st Sem (credited/ enrolled/ evaluations/approved, grade/without evaluations)	The number of curricular units credited/enrolled/evaluations/approved/ grade/ without evaluations by the student in the first semester.	Numerical
	Curricular unit's 2st Sem (credited/ enrolled/ evaluations/approved, grade/without evaluations)	The number of curricular units credited/enrolled/evaluations/ approved/grade/ without evaluations by the student in the second semester.	Numerical
	Application mode	The method of application used by the student	Categorical
	Application order	The order in which the student applied	Numerical
Social-economic Data	Mother's/Father's qualification /occupation	The qualification/occupation of the student's mother/father	Categorical
	Tuition fees up to date	Whether the student's tuition fees are up to date	Categorical
	Scholarship holder	Whether the student is a scholarship holder	Categorical
	Educational special needs	Whether the student has any special educational needs	Categorical
	Debtor	Whether the student is a debtor.	Categorical
	Unemployment rate	Unemployment rate among students	Numerical
	Inflation rate	Inflation rate of the region	Numerical
	GDP	Gross domestic product (GDP) from the region	Numerical

C. Data Preparation

Data preprocessing is a crucial step in knowledge discovery, involving cleaning, reduction, transformation, and feature selection. It involves inspecting the dataset for unwanted values, eliminating irrelevant fields, and converting the cleaned data into a suitable format for effective machine learning algorithms [16]. During this step, tasks such as checking missing data and duplicating data, replacing numerical values with names, changing categorical columns to categorical data type, getting all categorical variables except target, and rename columns were carried out.

D. Modeling

Previous studies showed that data mining involved creating models for classification, prediction, or finding hidden patterns in data that was observed [9, 10, 11]. The data mining that is prevalent these days can be divided into two types: the supervised and unsupervised data mining. However, supervised algorithms work on the basis of learning the category of unseen data, while unsupervised algorithms enable the study of the hidden patterns without target variable [19]. Logistic Regression (LR), XGBoost (XGB), Random Forest (RF), Support Vector Machines (SVM), and CatBoost (CB) classifiers were utilized as the supervised approaches to develop the models for this study.

1) Logistic regression: One prevalent linear classification approach that is effective at simulating the likelihood of a binary result of 1 or 0 is a logistic regression; it is either primarily dependent on one or more input factors. The dependent variable is a generalized combination of the predictor variables which is expressed as the logistic regression equation modeled by the logistic function, also known as the logistic or sigmoid function [21].

2) XGBoost (XGB): XGB is a supervised method which heavily relies on a specific algorithm called boosting. Starting with a learning set which includes correctly labeled samples, it is brought to a model of prediction which is used on any new examples. The training example pairs (x_0, y_0), (x_1, y_1) ... (x_n, y_n) are used by the method, where x is a vector of features and y is its label. Regression and classification issues where the label y takes a discrete or continuous value can be solved with it. The decision tree boosting technique XGBoost develops many models in a sequential fashion, each one aiming to make up for the shortcomings of the one before it. Hence, by using the XGBoost algorithm, this method can be expanded to a generalized gradient boosting [22].

3) Random Forest (RF): A Random Forest classifier is utilized for the regression and classification tasks, which involves a specific machine learning technique. It improves the accuracy by creating decision trees that are trainable on different parts of the same training set. Thus, to prevent overfitting, it learns from all the parts of the data. Random Forest generates K numbers of trees with distinct attributes each time, without pruning, by selecting attributes at random. Besides, test data is examined on every created tree, as opposed to decision trees, and the most frequent output is assigned to that particular instance [23] A forest with a larger number of trees yields the best accuracy. Missing values and category values are all handled by Random Forest. It gauges the purity and impurity of attributes using the Gini index indicator. In general, Random Forest classifiers exhibit more robustness and efficiency in comparison to decision trees [24].

4) Support Vector Machine (SVM): Vapnik's theoretical learning concept serves as the foundation for SVM. Systemic risk minimization is embodied by SVM [21]. SVMs are used in various domains related to outlier detection, regression, and classification. An SVM's original input space is mapped into a high-dimensional dot product space via a kernel. The new area is known as the feature space, and it is there that the best hyperplane for maximizing generalizability is identified. A small set of data points known as support vectors can determine the ideal hyperplane. Even though it lacks problem-domain knowledge, an SVM can produce high generalization results for classification tasks [25].

5) CatBoost(CB): A machine learning technique called CatBoost Classifier (CBC) helps in regression, classification, multi-class classification, and ranking. The objective function is reduced as the gradient decreases, and this causes variations in its design. Furthermore, built-in analytics in CatBoost evaluate the accuracy of the model before deployment. It does away with the necessity for feature processing by introducing a novel technique for managing category attributes. In addition, CatBoost makes use of ordered boosting, a permutation-driven substitute for traditional boosting techniques. Unlike the Gradient Boost, which tends to overfit quickly, it contains a mod to handle overfitting in small datasets. Additionally, CatBoost has two boosting modes: Plain and Ordered [26].

A computer with 8GB of RAM and 11th Generation Intel Core i5 processor was used for the research. Moreover, Google Colab platform was used during the entire project's development. Pandas, NumPy, Matplotlib, Seaborn, and Scikit-Learn were among the toolkits used to evaluate and compare the suggested classification models.

In addition, the dataset was manipulated in a manner that produced a 10-fold cross-validation technique which was used in splitting the dataset in training and testing subsets. Also, the stratified K-Fold subfunction for the cross-validation was employed. Scikit-Learn's model selection function [19] was used to carry out this division. The evaluations on the performance scores of classifiers were done using the cross-validation score and the GridSearchCV subfunctions.

E. Evaluation

Performance of the models are evaluated using a number of metrics, such as accuracy, precision, F1-score and recall [27]. Following metrics are used for evaluating models.

1) Precision: It is defined as the ratio of the model's total predictions to the number of true positives (TP).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

2) Recall: It is computed by dividing the total number of false negatives (FN) and true positives (TP) by the number of true positives.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

3) Accuracy: It can be calculated by dividing the total number of correctly classified examples by the total number of classified examples.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

4) F1-Score: The precision and recall score averaged with respect to their weights is the F1 score. It also has consideration of false positives and fales negatives for it is to accurately indicate accuracy as well as recall.

$$\text{F1-Score} = \frac{2 * Recall * Precision}{Recall + Precision} \quad (4)$$

TP denotes true positive: The predicted category is positive and the actual category is positive. FP designates false positive: The predicted category is positive and the actual category is negative. FN represents false negative: The predicted category is negative and the actual category is positive. TN means true negative. The predicted category is negative and the actual category is negative.

F. Deployment

The system predicted student performance such as "graduate, dropout, and enrolled" and suggested helpful strategies for learning to support them. User input features like demographic, academic data, and socio-economic data can be entered into the system.

The study utilized Python as the main programming language, Pandas for data manipulation and preprocessing, and NumPy for numerical operations. Machine learning library like Scikit-Learn was used for implementing models. Matplotlib and Seaborn were used for visualizing data distributions and performance metrics. The models were trained on an 11th Generation Intel Core i5 processor, 8GB RAM, and Google Colab notebook platform. The final recommendation system was developed and deployed using Streamlit [28], a Python-based framework for building and deploying machine learning models as web applications. The web interface allowed users to upload student data and recommend personalized learning strategies based on input.

Moreover, this research utilized some recommended learning strategies in educational psychology as follows [15].

1) Learning strategies for dropout students: The students opt to that they could dropping out should be enrolled in a soft skills program such as time management, self-testing, note-taking, and other productive study methods. Additionally, the program should be supplemented with academic counseling and follow-up to help the students consistently integrate the techniques they have learned. It is also advisable to boost active-learning approaches including creating notes and rewriting them, practicing self-testing, but not relying on other activities in passive reading. This course further should support the students in both increasing their efficiency in pursuing a reasonable time schedule and trying to devote the specified amount of time required for learning every week. Lastly, it should address any motivational problems or negative attitudes towards difficult material that may cause students to lose interest.

2) Learning strategies for enrolled students: The evaluation of students who are new to the learning system is important. The skills the students have must be evaluated and if there are any weakness, such should be identified. Once the system has provided active learning techniques, created a schedule, dealt with the obstructs, advised on the best practices and offered attention and accountability good results can be expected.

3) Learning strategies for students likely to graduate: For students who are positively engaging with learning, the system should suggest time management, note-taking, active reading, and self-testing techniques. Furthermore, it should motivate them to improve advanced skills like research, academic writing, and critical thinking.

IV. Results and Evaluation

A. Descriptive Analysis

Firstly, the influence of demographic, socioeconomic, and academic data on student performance in higher education was investigated.

1) Student performance: Enrol, Graduate, and Dropouts : The pie chart of Fig. 3 displays three segments representing student performance in higher education: Dropout (50%), Graduate (32%), and Enrolled (18%). Such a chart was used to focus on parameters of student retention and graduation rates. This may be part of a discussion with broader goals including the effectiveness of academic support systems.

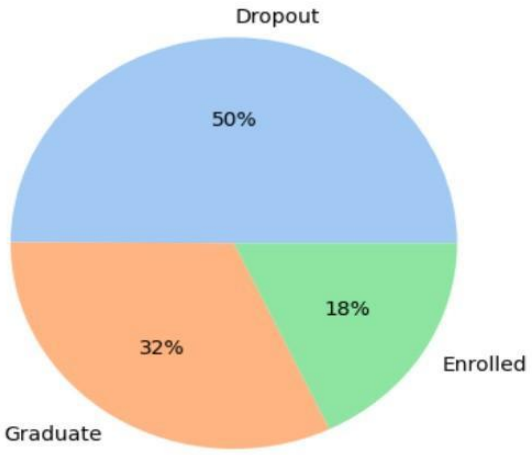


Fig. 3. Pie chart of student outcomes.

2) Top 10 factors that affect student performance: The heatmap in Fig. 4 illustrates the 10 most significant factors that affect student performance. The result of the heatmap visualized factors range from academic factors such as grades

of curricular unit and approvals to socio-economic factors such as scholarships and tuition fees.

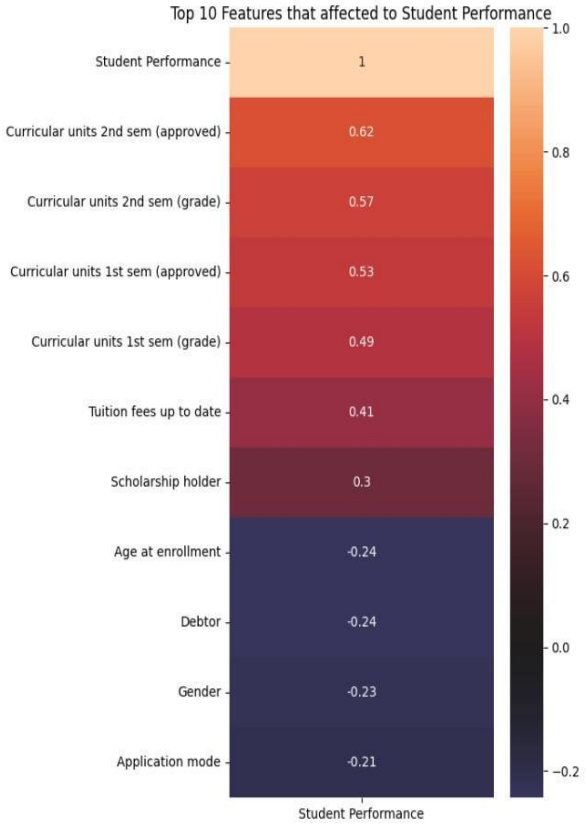


Fig. 4. Top 10 most influencing factors.

3) Age Distribution at the time of Enrollment: The histogram in Fig. 5 shows the age distribution of the students at the time of enrollment. The majority of enrolled students fell into the 18–22 age range, but it was evident that students in the 60–70 age demographic range are marginally present in the population.

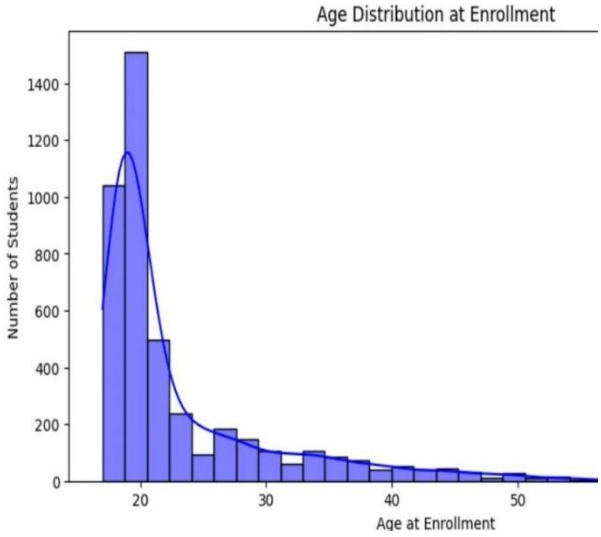


Fig. 5. Histogram of age distribution at the enrollment.

4) Effect of curricular units 2nd semester (approved) on student performance: The Fig. 6 bar chart exhibits a clear upward trend from dropouts to graduates in total curricular unit of 2nd semester (approved).

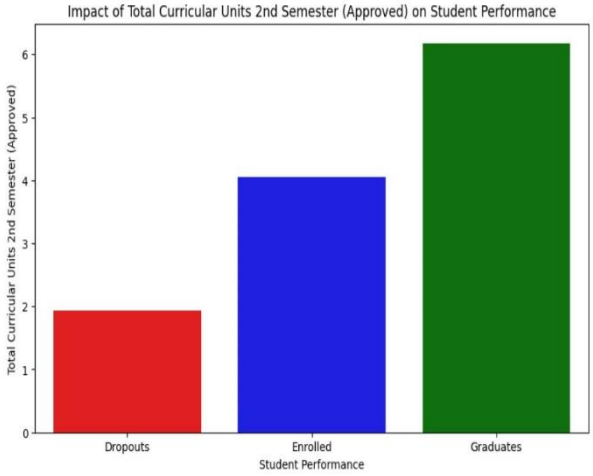


Fig. 6. Bar chart of student performance in all curricular units during the 2nd semester.

5) Effect of scholarship status on student performance: The stacked bar chart in Fig. 7 clearly illustrates a majority of scholarship holders graduating when it compared to non-scholarship holders.

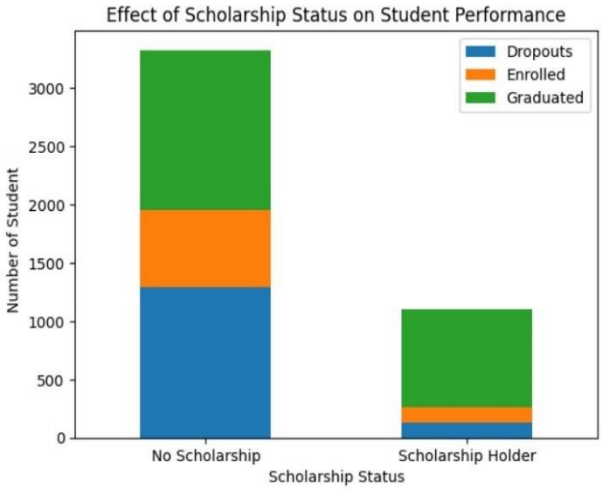


Fig. 7. Stacked bar chart of scholarship status on student performance.

B. Feature Selection

Secondly, a data mining method (classification) was applied to predict the performance of students based on affective factors. To identify the best prediction model, five distinct classifiers—Random Forest (RF), Support Vector Machine (SVM), XGBoost (XGB), CatBoost (CB), and Logistic Regression—were employed in the analysis. Each classifier was methodically assessed to determine the best outcome. Consequently, accuracy, F1-score, precision, and recall were measured, with the best results boldfaced. The results of the assessment of the selected classifiers are summarized and presented in Table II.

TABLE II. FEATURE SELECTION USING CLASSIFICATION TECHNIQUES

Algorithms	Accuracy	F1 - Score	Precision	Recall
Random Forest - RF	**0.7989**	0.7857	**0.7883**	**0.7989**
Support Vector Machine - SVM	0.7740	0.7690	0.7693	0.7740
XGBoost - XGB	0.7921	**0.7887**	**0.7883**	0.7921
CatBoost - CB	0.7955	0.7859	0.7837	0.7955
Logistic Regression - LR	0.7831	0.7686	0.7663	0.7831

With an accuracy of 79.89%, the Random Forest Classifier emerged as the leader and was now tied with the XGBoost Classifier for the highest F1-score of 78.87%. Based on these two metrics, the Random Forest Classifier was the best-performing model on the list. However, given the identical F1-scores, the XGBoost Classifier remain a formidable contender. The Random Forest Classifier is a robust and efficient method for addressing overfitting and noisy data in diverse and complex educational datasets [23]. It employs an ensemble learning approach, combining multiple decision trees, making it well-suited for handling missing values and categorical variables without extensive preprocessing, common challenges in educational data.

C. Recommendation System

Finally, a recommendation system was developed that uses a pre-trained Random Forest model to predict student performance, categorizing it into three probable statuses: enrolled, graduate, and dropout.

Based on predictions, the system delivers different learning strategy recommendations. The following are the steps in the system: Upload a csv file or manually input features, Load the pre-trained Random Forest model, predict the outcome using a model, and Display prediction probabilities and recommendations.

1) User interface: Fig. 8 depicts the web interface to get input features.

2) The Student is likely to graduate: Three categories display the prediction probabilities: 13.46%, 15.82%, and 70.72% for graduates, enrolled, and dropouts, respectively. The predicted probabilities suggested that the student had a positive chance of graduating and offered suggested learning techniques. These details are displayed in Fig. 9.

3) The Student is likely to drop out: The prediction probabilities indicate that there is a significantly larger probability dropping out at 59.80%, enrolling at 10.51%, and graduating at 29.69%. It is shown in Fig. 10.

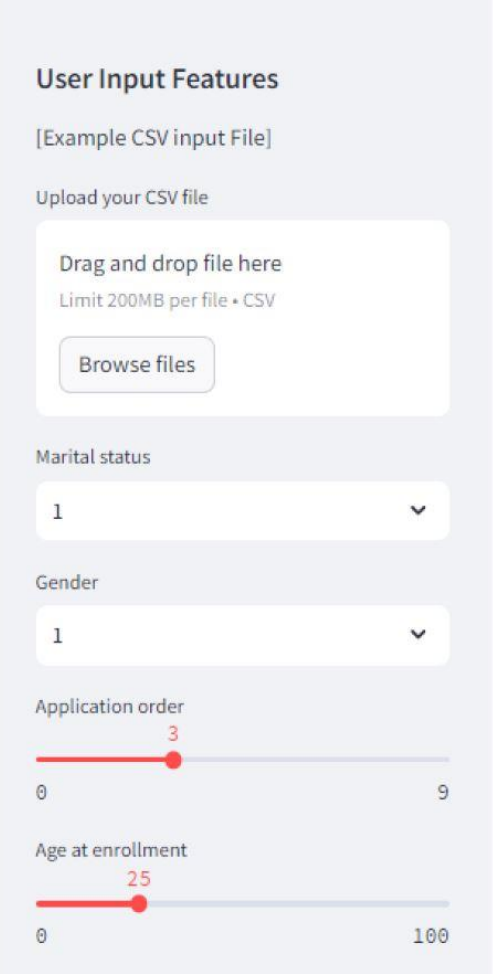


Fig. 8. Web interface to get input features.

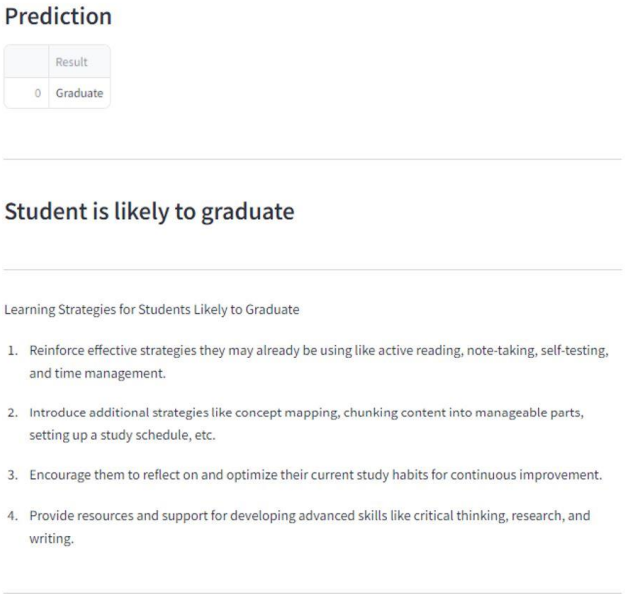


Fig. 9. Prediction probabilities and learning strategies of student is likely to graduate.

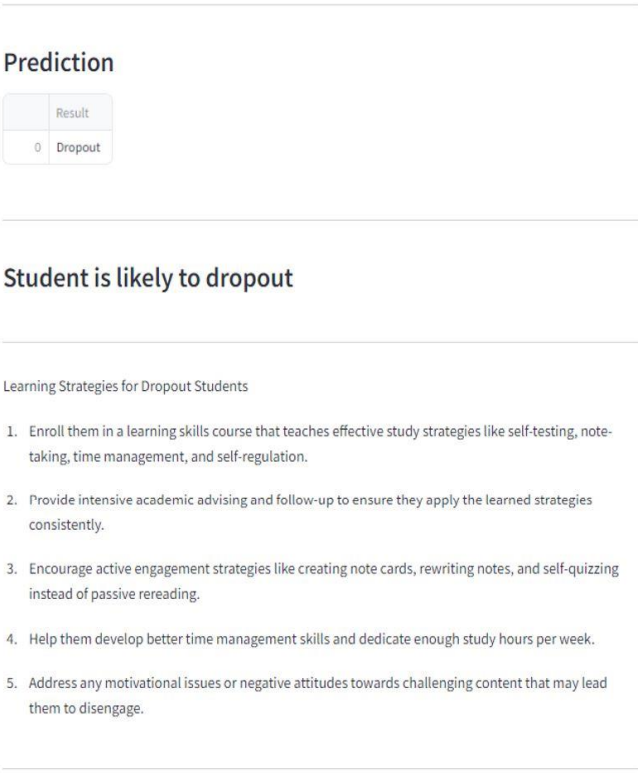


Fig. 10. Prediction probabilities and learning strategies of student is likely to dropout.

V. Discussion

In this section, the findings from the descriptive analysis, feature selection, and the recommendation system presented in the previous section are discussed.

A. The Influence of Demographic, Socioeconomic, and Academic Factors

This article generally indicates the vital part that demographic, socioeconomic and academic factors play in the tertiary education performance of the students. Contrary to the graduates (32%) and those who are still enrolled (18%), the very high dropout rate (50%) demonstrated that higher education system has a serious issue. This distribution in fact suggests that outside factors that might be related to socioeconomic status and demographic backgrounds, play a significant role in the outcomes of learners even when there are educational programs.

The factors that affect student performance further support the complex relationship between student success and socio-economic factors, such as scholarships, tuition costs, and academic indicators like the number of curricular units in semesters. This implies a need for assistance with both academic challenges and economic hurdles. Consequently, the results of this research are also supported by the findings of [9, 13]. In Song et al. [9], the authors explored issues including behavior, academics, and demographics. Behavioral components englobed being absent from class, while academic achievement, coupled with good grades was fundamentally important. Socioeconomic status and year of entry were the demographic elements involved. Even during Li et al. [13], the authors went for the details of demographic variants like age, occupations of parents, and nationality, which connected the rich diversity of students with the advantages and challenges of different cultures. The socioeconomic factors, like parents' level of education or income of the family, influenced their education, the availability of resources and their academic success to a great extent. Academic factors, one of which is previous performance and absences, are directly correlated with the student performance and therefore, they have a chance to predict future outcomes.

B. Efficiency of Classification Methods in Predicting Student Performance

Applications of many classifiers, the obtaining of insightful results, and the prediction of student achievement have been done. The Random Forest classifier [23] has an accuracy of approximately 80%; it performs better than other models, and what makes it robust is its ability to efficiently handle many different kinds of datasets that are common in educational settings. Random Forest's performance is an indication that such ensemble learning approaches are essential for successful educational data mining. These results are invaluable for developing algorithms that can identify at-risk student cohorts, enabling timely interventions.

The study [10] employs different types of classification for the prediction of student success and the grade of assessment is revealed to be the major factor in rates of performance. The models that were using the grades and online activity data performed better. However, it was the Random Forest classification model that performed better in predicting student performance.

C. Implications of Recommendation System

The development of a recommendation system that classifies students to likely categories (graduate, enrolled, or dropout) and provides specific educational strategies based on the predictions is especially inventive. This technology is very promising, it puts machine learning and its predictive power to real use in a way that has a practical impact on the education process, and can completely change the concept of teaching and learning. E.g., 59.80% dropout rate of this student could trigger the implementation of special supportive measures like tutoring, counseling, or modification of financial aids by the academic institutions.

In addition, a similar conclusion was reached in studies [14, 15], which emphasize the significance of using learning analytics with AI-based performance prediction models. The results of such a technique demonstrate that cooperation is enhanced, that satisfaction is higher and that engagement and learning are improved. Such study shows the trend toward educational effectiveness by emphasizing the application of AI models which have been created and the gap bridging between the creation and use of AI models.

D. Implications and Limitations

By personalizing progressions of study to suit the needs of individual students, offering predictive insights for early intervention and continuously enhancing instructional methodologies and techniques, AI can improve personalized

learning [15]. It could thus lead to improved academic work, a drop in cases of dropout rate as well as more effective learning.

The amount, class, and constant quality of data, which are available, determine how efficient AI systems work [29]. They might tarnish the trust because the terms are hard to be understood or difficult to comprehend. Differing from one to the other in practices and demography could limit the applicability of the same interventions in various settings [14]. In fact, information security and privacy are among the dilemmas that ethical corporations often encounter [30]. If AI extensively used, task automation will supplant people interaction and people will experience threats of a decline in critical thinking and social skills. The issue of resistance, a lack of technical compatibility, and the need for educating the staff may complicate automated systems integration with existing educational infrastructure.

VI. Conclusion

This study titled "EDM and AI-Powered Learning Strategy Recommendation System to Boost the Academic Results" articulates that AI and EDM can be used in learning settings to improve academic performance. Students' results prediction model and recommendation engine which use variables like socioeconomic status, demographic data source and academic performance was built using the CRISP-DM model. Notably, in this case, this research reconciled this navigation with machine learning algorithms in a perfect way by proving that random forest was the ideal classifier for the student's performance status prediction. With the implied procedure proposed as a basis, unique study systems were built for the learners on the foundation of their particular needs, desires, and struggles. The fact that personalized education might be achieved within these systems is a crucial milestone for the field of personalized learning in general. The evidence has a significant impact that will aid institutes to formulate student progress and knowledge exploration concerning factors that invite student success. In this regard, the AI-powered system which would sense learners and respond with personalized instructions should promote the students' access rights, involvement, and productivity. Through this shift, the numbers of student dropout might come down and a comfortable atmosphere in which the learning process can be done effectively and this could be realized.

In the future, real-time data analytics integration and ongoing AI algorithm improvement should improve on assessing the productivity and adaptability of learning systems. Moreover, investigating hybrid models that fuse AI suggestions with the knowledge of human educators may result in more comprehensive teaching approaches.

References

[1] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art ," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 6, pp. 601-618, Nov. 2010, doi: 10.1109/TSMCC.2010

[2] C. Romero and S. Ventura, "Data mining in education," WIREs Data Mining and Knowledge Discovery, vol. 3, pp. 12-27, 2012, doi: 10.1002/widm.1075

[3] A. Dutt, M. A. Ismail and T. Herawan, "A Systematic Review on Educational Data Mining," in *IEEE Access*, vol. 5, pp. 15991-16005, 2017, doi: 10.1109/ACCESS.2017.2654247

[4] B. K. Francis and S. S. Babu, "Predicting academic performance of students using a hybrid data mining approach," Journal of Medical Systems, vol. 43, no. 6, Apr. 2019. doi:10.1007/s10916-019-1295-4.

[5] D. Hooshyar, M. Pedaste, and Y. Yang, "Mining educational data to predict students' performance through procrastination behavior," *Entropy*, vol. 22, no. 1, p. 12, Dec. 2019. doi:10.3390/e22010012

[6] T. Liu, C. Wang, L. Chang, and T. Gu, "Predicting high-risk students using learning behavior," *Mathematics*, vol. 10, no. 14, p. 2483, Jul. 2022. doi:10.3390/math10142483

[7] N. I. Mohd Talib, N. A. Abd Majid, and S. Sahran, "Identification of student behavioral patterns in higher education using K-means clustering and support vector machine," *Applied Sciences*, vol. 13, no. 5, p. 3267, Mar. 2023. doi:10.3390/app13053267

[8] T. Tao, C. Sun, Z. Wu, J. Yang, and J. Wang, "Deep neural network-based prediction and early warning of student grades and recommendations for similar learning approaches," *Applied Sciences*, vol. 12, no. 15, p. 7733, Aug. 2022. doi:10.3390/app12157733

[9] Z. Song, S.-H. Sung, D.-M. Park, and B.-K. Park, "All-Year Dropout Prediction Modeling and Analysis for University Students," Applied Sciences, vol. 13, no. 2, p. 1143, Jan. 2023, doi: 10.3390/app13021143.

[10] A. Alhassan, B. Zafar, and A. Mueen, "Predict students' academic performance based on their assessment grades and online activity data," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 4, 2020. doi:10.14569/ijacsa.2020.0110425

[11] A. F. Mohamed Nafuri, N. S. Sani, N. F. A. Zainudin, A. H. A. Rahman, and M. Aliff, "Clustering Analysis for Classifying Student Academic Performance in Higher Education," Applied Sciences, vol. 12, no. 19, p. 9467, Sep. 2022, doi:10.3390/app12199467.

[12] E. M. Queiroga et al., "Using Virtual Learning Environment Data for the Development of Institutional Educational Policies," Applied Sciences, vol. 11, no. 15, p. 6811, Jul. 2021, doi: 10.3390/app11156811.

[13] F. Li, Y. Zhang, M. Chen, and K. Gao, "Which Factors Have the Greatest Impact on Student's Performance," Journal of Physics: Conference Series, vol. 1288, p. 012077, Aug. 2019, doi: 10.1088/1742-6596/1288/1/012077.

[14] F. Ouyang, M. Wu, L. Zheng, L. Zhang, and P. Jiao, "Integration of Artificial Intelligence Performance Prediction and learning analytics to improve student learning in online engineering course," *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, Jan. 2023. doi:10.1186/s41239-022-00372-4

[15] S. J. Renzulli, "Using learning strategies to improve the academic performance of university students on academic probation," NACADA Journal, vol. 35, no. 1, pp. 29–41, Jul. 2015. doi:10.12930/nacada-13-043

[16] C. Shearer, "The CRISP-DM model: The new blueprint for data mining," J. Data Warehousing., vol. 5, no. 4, pp. 13–22, 2000.

[17] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISPDM process model," Procedia Comput. Sci., vol. 181, pp. 526–534, 2021. doi:10.1016/j.procs.2021.01.199

[18] W. F. Yaacob, S. A. Nasir, W. F. Yaacob, and N. M. Sobri, "Supervised Data Mining Approach for predicting student performance," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 16, no. 3, p. 1584, Dec. 2019. doi:10.11591/ijeecs.v16.i3. pp1584-1592

[19] M. Bellaj, A. Ben Dahmane, S. Boudra, and M. Lamarti Sefian, "Educational Data Mining: Employing Machine Learning techniques and hyperparameter optimization to improve students' academic performance," International Journal of Online and Biomedical Engineering (iJOE), vol. 20, no. 03, pp. 55–74, Feb. 2024. doi:10.3991/ijoe.v20i03.46287

[20] "Predict students' dropout and academic success," Kaggle, https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention (accessed Apr. 18, 2024).

[21] A. Salah Hashim, W. Akeel Awadh, and A. Khalaf Hamoud, "Student performance prediction model based on supervised machine learning algorithms," *IOP Conference Series: Materials Science and Engineering*, vol. 928, no. 3, p. 032019, Nov. 2020. doi:10.1088/1757-899x/928/3/032019

[22] R. Mitchell and E. Frank, "Accelerating the XGBOOST algorithm using GPU computing," *PeerJ Computer Science*, vol. 3, Jul. 2017. doi:10.7717/peerj-cs.127.

[23] N. S. Ahmed and M. Hikmat Sadiq, "Clarify of the random forest algorithm in an educational field," *2018 International Conference on Advanced Science and Engineering (ICOASE)*, Oct. 2018. doi:10.1109/icoase.2018.8548804

[24] M. M. Tamada, R. Giusti, and J. F. Netto, "Predicting students at risk of dropout in technical course using LMS Logs," *Electronics*, vol. 11, no. 3, p. 468, Feb. 2022. doi:10.3390/electronics11030468

[25] Y. Chen, "Support Vector Machines and Fuzzy Systems," *Soft Computing for Knowledge Discovery and Data Mining*, pp. 205–223, 2008. doi:10.1007/978-0-387-69935-6_9

[26] A. Joshi *et al*., "CatBoost — an ensemble machine learning model for prediction and classification of student academic performance," *Advances in Data Science and Adaptive Analysis*, vol. 13, no. 03n04, Jul. 2021. doi:10.1142/s2424922x21410023

[27] G. Feng, M. Fan, and C. Ao, "Exploration and visualization of learning behavior patterns from the perspective of educational process mining," *IEEE Access*, vol. 10, pp. 65271–65283, 2022. doi:10.1109/access.2022.3184111

[28] M. Khorasani, M. Abdou, and J. Hernández Fernández, "Streamlit Basics," *Web Application Development with Streamlit*, pp. 31–62, 2022. doi:10.1007/978-1-4842-8111-6_2

[29] M. Murtaza, Y. Ahmed, J. A. Shamsi, F. Sherwani, and M. Usman, "AI-based personalized e-learning systems: Issues, challenges, and solutions," IEEE Access, vol. 10, pp. 81323–81342, 2022. doi:10.1109/access.2022.3193938

[30] K. Zhang and A. B. Aslan, "AI Technologies for Education: Recent research & future directions," Computers and Education: Artificial Intelligence, vol. 2, p. 100025, 2021. doi:10.1016/j.caeai.2021.100025

Implementation of Lattice Theory into the TLS to Ensure Secure Traffic Transmission in IP Networks Based on IP PBX Asterisk

Olga Abramkina[1], Mubarak Yakubova[2], Tansaule Serikov[3], Yenlik Begimbayeva[4], Bakhodyr Yakubov[5]
Department of Cybersecurity, International Information Technology University, Almaty, Kazakhstan[1]
Department of Cybersecurity, Almaty University of Power Engineering and Telecommunications
Name after Gumarbek Daukeev, Almaty, Kazakhstan[2, 4, 5]
Department of Electronics and Telecommunication, S.Seifullin Kazakh AgroTechnical Research University, Astana, Kazakhstan[3]

***Abstract*—This paper presents a novel lattice-based cryptography implementation in the Transport Layer Security (TLS) protocol to enhance the security of traffic transmission in IP networks that use the Asterisk IP PBX platform. Given the growing threat of quantum computing, traditional cryptographic methods are becoming increasingly vulnerable. To address this issue, the study leverages post-quantum cryptography by developing a modified TLS protocol using lattice-based cryptographic algorithms. The performance of the system was evaluated in terms of security, computational efficiency, and real-time communication. The study shows that the proposed lattice-based TLS implementation effectively secures traffic transmission in IP PBX networks, offering a robust solution against both current and future quantum threats.**

Keywords*—IP; PBX; Asterisk; TLS; MITM; post-quantum cryptography

I. Introduction

In today's rapidly advancing technological world, the prospect of a quantum computer is becoming increasingly real. Quantum computers, once fully implemented, could potentially break widely used cryptographic algorithms such as RSA and ECC (elliptic curve cryptography) [1-3].

If the intruder's computing capabilities increase by tens, hundreds or thousands of times, this will lead to a sharp need to increase the key length to a critical level, which will become unsuitable for successful operation in real information systems. In addition, if a quantum adversary with enormous computing power appears, there is a possibility of a complete hack of existing cryptosystems by a complete enumeration of keys [4-6]. This problem can be solved by implementing post-quantum cryptography, a relatively new area of cryptography, which is designed to resist quantum computing [7-9]. Unlike asymmetric cryptography, based on conditionally unidirectional mathematical functions, post-quantum cryptography is based on the principles of quantum mechanics and quantum information theory, which guarantee physical unidirectionality. However, there are problems associated with the complexity of implementation and the high cost of equipment [10-12]. With a data transmission channel length of more than 100 km, the transmission speed is significantly reduced (to several bits per second). This fact does not yet allow for the implementation of a full-fledged secure exchange of critical information. In this regard, post-quantum cryptography currently appears more feasible for use in existing systems. The main approaches in post-quantum cryptography include:

1) Lattice-based cryptography. This approach is based on mathematical problems involving lattices, such as the shortest vector problem (SVP) and learning with errors (LWE). These problems are considered difficult to solve even for quantum computers, so lattice cryptography has an advantage [13-15].

2) Code-based cryptography. This approach is based on the difficulty of decoding a general linear code, a problem that has been studied extensively and is considered difficult for quantum computers [16-18].

3) Multivariate polynomial cryptography. This approach is based on the difficulty of solving systems of multivariate quadratic equations over a finite field. The multivariate quadratic (MQ) problem is NP-hard and is considered resistant to quantum attacks [19-21].

4) Hash-based cryptography. This approach uses cryptographic hash functions as a basis for constructing secure algorithms, particularly for digital signatures. Hash-based cryptography is widely considered to be secure against quantum attacks, assuming that the underlying hash functions are secure [22-25].

5) Isogeny-based cryptography. This approach exploits the mathematical properties of isogenies between elliptic curves. Isogeny-based cryptographic schemes are relatively new and are considered promising due to their resistance to quantum attacks. Small key sizes and robust security based on complex mathematical structures [26-28].

6) Hyperelliptic curve cryptography uses the algebraic structures of hyperelliptic curves, which generalize elliptic curves and provide similar security mechanisms at smaller key sizes [29-31]

7) Symmetric quantum resistance focuses on making symmetric cryptosystems secure against quantum computing threats. The primary method of improving the resistance is to increase the key length and choose algorithms with greater resistance to Grover-based attacks. For example, using AES-256 instead of AES-128 will provide resistance against

quantum computing attacks at the level of 2128 searches. Quantum attacks on hash functions, such as the Grover-based collision attack, also reduce the complexity of finding a collision from O(2n)) to O(2n/2). Therefore, using hash functions with longer bit lengths (e.g., SHA-512 instead of SHA-256) can be an effective means of protection [32-35].

These approaches form the basis of ongoing research and development in post-quantum cryptography as the world prepares for the potential impact of quantum computing. Each approach has its own strengths and weaknesses, and the future of secure communications will likely involve a combination of these methods.

This paper is a continuation of the research [36] where the publication focused on traditional multi-layered data protection in IP networks based on Asterisk IP PBX using different codecs to reduce latency and improve real-time encryption, while this study proposes the integration of post-quantum lattice-based cryptography into the TLS protocol to protect against threats associated with the development of quantum computers. The [36] solves the existing problems of cryptographic security, and this study extends the research area by proposing protection against future quantum threats, which represents an important step forward in ensuring the security of IP networks.

This article discusses the convergence of lattice theory into the TLS protocol for protecting transmitted traffic in IP networks based on Asterisk PBX. Asterisk is an open source framework for building communications applications such as IP PBX systems, VoIP gateways, and conference servers. Given its widespread use, ensuring the security of communications processed by Asterisk is of paramount importance.

To address these issues, lattice-based cryptography has become a promising approach to post-quantum cryptography, offering strong resistance to quantum attacks.

Over the past decade, the number of published works on this topic, both foreign and domestic, has increased significantly. This fact emphasizes the relevance of this problem and arouses interest in further research in this area.

II. Methodology

In the modern conditions of quantum computing development, traditional cryptographic methods based on the complexity of factorization and discrete logarithm calculation become vulnerable to quantum attacks. Therefore, there is a need to use post-quantum cryptographic algorithms that will ensure security even in the conditions of powerful quantum computing. Lattice-based Cryptography is considered one of the most promising technologies in the field of post-quantum cryptography. The essence of the approach is based on the difficulty of solving problems on lattices, such as the problem of learning with errors (LWE), which remains difficult even for quantum computers [37-40].

These methods were replaced by a lattice algorithm. The LWE problem was proposed as an approximate version of the Shortest Vector Problem (SVP) on lattices. The difficulty of LWE is directly related to solving lattice problems such as the Closest Vector Problem (CVP), since finding a secret vector s in the LWE problem reduces to solving a lattice variation of the CVP problem if the matrix A represents a lattice basis. One can say that LWE translates lattice problems into an algebraic form, where noise (error) is added to a system of linear equations.

The proposed methodology introduces the use of lattice-based cryptographic schemes to provide post-quantum security in the process of key exchange and data encryption. The key exchange scheme is based on LWE or the related Ring-LWE approach, which provides a high level of resistance to both classical and quantum attacks.To integrate such an algorithm into Asterisk, an additional module was created. This module is responsible for generating and verifying keys. Third-party cryptographic libraries such as liboqs (Open Quantum Safe), which supports various post-quantum algorithms, are integrated via AGI. The Asterisk dialplan is then configured to call these modules when establishing a secure SIP session.

Modified TLS (with post-quantum cryptography) has the following implementation stages:

1) The client requests a secure connection.

2) The server sends its certificate based on the post-quantum PQC algorithm.

3) The client and server exchange keys using the PQC algorithm.

4) The symmetric key is used to encrypt data using AES.

5) Messages are protected using HMAC, as in classic TLS, but the key itself was transmitted using PQC.

To generate keys according to the PQC algorithm, a random matrix A and a vector s are used. The matrix A belongs to the set Zq^m: where q is a prime number.

The vector s belongs to the set Zq^m: and is the secret key.

The public key is formed from the matrix A and the error vector e.

The vector b = As + e is calculated, where e is a random noise vector.

b and A make up the public key, and s makes up the secret key.

The encryption process is described by calculating the components of the ciphertext from the vectors u, v, after generating a random vector r from the set Zq^m:

$$u=A^Tr, \quad (1)$$

where, u is the portion of the ciphertext associated with the public key.

$$v=b^Tr+m\cdot\lfloor 2/q \rfloor, \quad (2)$$

where, v contains the message m itself with added noise.

During decryption, the original message is calculated:

$$m=decode(v-uTs), \quad (3)$$

where, the expression v−uTs allows us to obtain the encrypted message, and then decode it back into the original message.

The module code is shown in Fig. 1.

```
import oqs
import socket
import ssl
import logging
from queue import Queue
from collections import defaultdict
import secrets
from threading import Thread
import datetime
import sqlite3
from cryptography.hazmat.primitives.kdf.pbkdf2 import PBKDF2HMAC
from cryptography.hazmat.primitives import hashes
from cryptography.hazmat.backends import default_backend
import certbot.main
import time
import shutil
import os
import threading


logging.basicConfig(filename='server.log', level=logging.INFO, format='%(asctime)s - %(levelname)s - %(message)s')
server_log = logging.getLogger("server")
security_log = logging.getLogger("security")
security_log.setLevel(logging.INFO)
security_handler = logging.FileHandler("security.log")
security_formatter = logging.Formatter('%(asctime)s - %(levelname)s - %(message)s')
security_handler.setFormatter(security_formatter)
security_log.addHandler(security_handler)


MAX_REQUESTS_PER_MINUTE = 100
MAX_DATA_SIZE = 4096
BACKUP_DIR = '/path/to/backup/dir'
```

Fig. 1. Module code.

The performance and productivity of the developed method were evaluated by conducting a Man-in-the-Middle (MITM) attack, a well-known threat to VoIP systems. In MITM attacks, an attacker intercepts and manipulates real-time data, gaining unauthorized access to voice calls, modifying messages, and retrieving confidential credentials. In the context of VoIP, implementing TLS (Transport Layer Security) effectively protects signaling traffic, significantly reducing the risk of MITM attacks. By integrating a post-quantum key exchange algorithm into TLS, the developed method ensures robust encryption and data integrity, safeguarding communications from both conventional MITM and future quantum-based attacks.

This approach enhances the overall security of VoIP systems against evolving threats.Stages of implementing a MITM attack in a VoIP environment to test TLS:

1) Setting up a test environment includes preparing a virtual machine (VM) with an installed VoIP system based on Asterisk and interaction of virtual clients via SIP.

2) The creation of a MITM attack is based on the use of Ettercap and Bettercap tools, which are capable of intercepting traffic between VoIP clients.

3) Testing without TLS, when protocols transmit messages in clear text, which allows an attacker to easily intercept calls and gain access to data.

4) Implementation *of modified TLS* allows us to evaluate how TLS affects data interception.

5) Analysis of the results shows the results of the MITM attack with and without the use of modified TLS.

III. Result

A. Installing Asterisk

Asterisk installed on a virtual machine to emulate a VoIP server was done using Docker.

Docker is an open-source platform that allows you to optimize the management of development, testing and deployment of web applications. Docker is based on the packaging of programs (along with the environment and dependencies) into virtual blocks - containers.

Using Docker for IP PBX Asterisk is an effective way to deploy, manage, and scale an IP telephony system. Asterisk, being one of the most popular platforms for building VoIP (Voice over IP) solutions, can benefit greatly from using containerization via Docker. The PBX AS-IS architecture is shown in Fig. 2.

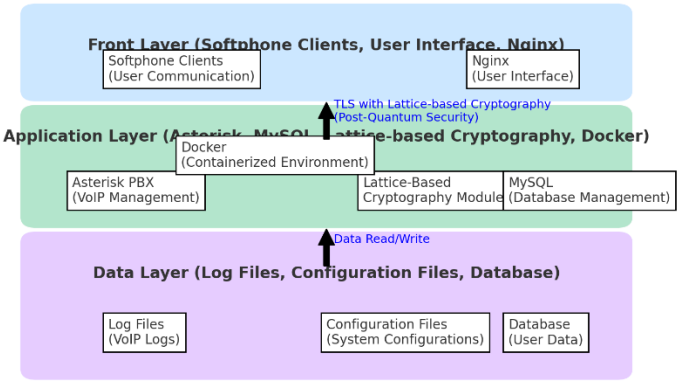


Fig. 2. Architecture of the proposed system.

In application architecture, especially with containerization and technologies like Docker, there are three key layers: Data Layer, Application Layer, and Front Layer. These layers provide separation of duties, simplify support, development, and scaling of the system.

1) Data layer: The data layer is responsible for storing, managing, and processing all the information used in the system. In the context of IP telephony architecture using Asterisk and Docker, this layer is responsible for storing and managing data such as call records, user settings, call sessions, and so on.

The main components of the data layer are:

Databases: Used to store configurations, user data, CDRs (Call Detail Records), logs, session information, and other necessary information.

MySQL, PostgreSQL, MariaDB: Relational databases for storing structured data about users, calls, and other elements of the system.

NoSQL databases (e.g. Redis): Used for caching, storing sessions or other data where high access speed is critical.

File storage: Can be used to record and store conversations, system logs and other files related to IP telephony.

Object Storage (e.g. MinIO or S3): For storing large volumes of audio files, call recordings, data archiving.

File systems (NFS, Ceph): For local or distributed data storage between different containers and system nodes.

Data caching: To optimize access to frequently used data and reduce the load on the main databases.

Redis or Memcached: Used to cache database queries or store session data in real time.

2) Application layer: The application layer is the main functional layer where the business logic of the system is executed. In the case of IP telephony and Asterisk, this is the layer where call processing, SIP request routing, security policy enforcement, etc. take place.

The main components of the application layer are:

Asterisk Server: The main component that handles voice traffic, SIP request routing, session management, and provides telephony functionality. It handles incoming and outgoing calls, controls conferences, organizes IVR systems, and other IP PBX functions.

SIP Proxy (e.g. Kamailio): A component for routing SIP traffic, load balancing between multiple Asterisk servers, and improving system security.

Security Features: Uses TLS to encrypt SIP messages and SRTP to protect voice traffic. These components are integrated into the main application server, providing protection against attacks such as MITM (Man-in-the-Middle).

Application business logic: Applications and services that provide telephone communication functions, such as IVR, call processing automation, integration with CRM systems, and others.

Additional services:

CDR (Call Detail Records): Systems for keeping track of calls and generating reports on the operation of the telephone network.

Call recording functions: For storing call records for security or archiving purposes.

Monitoring and logging: Services for tracking system performance, analyzing logs, and preventing errors.

3) Front layer: The front layer is responsible for user interaction with the system and providing them with interfaces for accessing functionality. These can be web interfaces, mobile applications, client programs for IP telephony, and other user interaction components.

Main components of the front layer:

IP telephony client applications: Programs that users use to work with the IP PBX system.

Softphone clients: For example, Blink, Zoiper, X-Lite are voice communication programs that work via the SIP protocol.

Web clients: Web interfaces for accessing Asterisk functions such as call management, user configuration, statistics analysis, etc.

Web management interfaces: Applications that allow administrators and users to interact with the system via web browsers.

Call control panels: Visual panels for managing calls in real time, viewing line status and interacting with the system.

Mobile applications: Programs for mobile devices that allow users to make and receive calls over a corporate telephone network using SIP.

API: Programming interfaces for integrating Asterisk with external systems (e.g. CRM, ERP). The API allows you to automate some processes and improve the functionality of the system for end users.

RESTful API or AMI (Asterisk Management Interface): Allows you to manage calls,Let's consider the main layers and elements of the architecture:

Installing Docker on Ubuntu distribution consists of 12 steps:

1) Updating the package: sudo apt update

2) Installing the package that is required for the apt package manager to work over HTTPS: sudo apt install apt-transport-https ca-certificates curl software-properties-common

3) Adding the Docker repository GPG key: curl -fsSL https://download.docker.com/linux/ubuntu/gpg | sudo apt-key add -

4) Adding the Docker repository: sudo add-apt-repository "deb [arch=amd64] https://download.docker.com/linux/ubuntu bionic stable"

5) Updating the package again: sudo apt update

6) Switching to the Docker repository to install it: apt-cache policy docker-ce

Similar information with the Docker version is shown in Fig. 3.

7) Install Docker: sudo apt install docker-ce

8) Check the program's functionality: sudo systemctl status docker

Information that Docker is active is shown in Fig. 4.

9) To use the docker utility, you need to add the user name to the Docker group: sudo usermod -a -G docker user

*10)*Enter the user name: su - user

*11)*Set the user password.

*12)*Check access to Docker images: docker run hello-world

After the information about the successful installation of "Hello from Docker!", you need to install PORTAINER - a graphical panel for managing docker containers.

A data storage for Portainer has been created:

docker volume create portainer_data

The container with Portainer is launched with the command:

docker run -d -p 9000:9000 -v /var/run/docker.sock:/var/run/docker.sock -v portainer_data:/data portainer/portainer

After launching in the browser at the server ip address:9000, you need to set the administrator password.

Next, select the location of Docker on the local server (Local) or on a remote one.

The panel is installed, you can launch containers (Fig. 5).

In the control panel, in the "App Template" section, you can find templates with software and run them in containers (Fig. 6).

Portainer is used for comfortable container management.

SIP clients are installed on software SIP phones (softphones).

B. *Conducting a MITM Attack*

Traffic interception is organized by setting up a network bridge and using an ARP spoofing tool (ettercap) so that the attacker's virtual machine can intercept traffic between SIP clients and the VoIP server. The command for an ARP spoofing attack:

ettercap -T -M arp:remote /client-IP-address/ /server-IP-address/

Wireshark was used to analyze traffic. Filters in Wireshark for VoIP:

sip || rtp

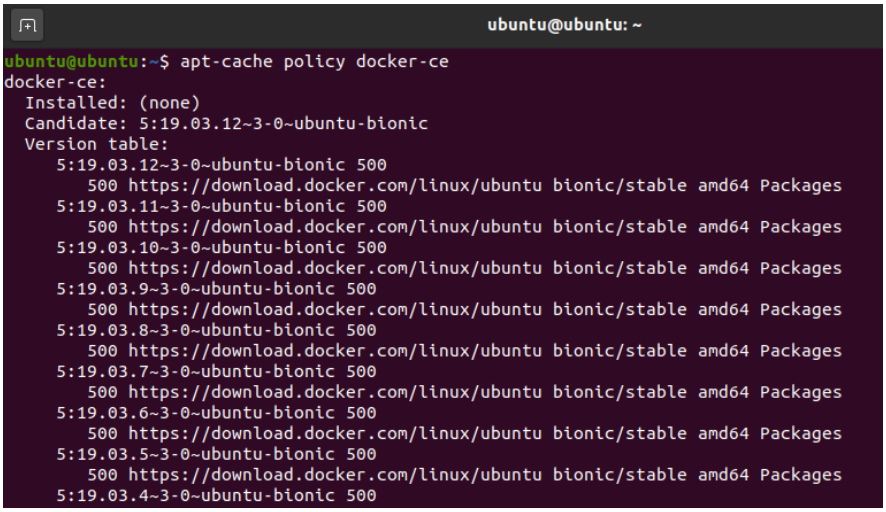


Fig. 3. Similar information with docker version.

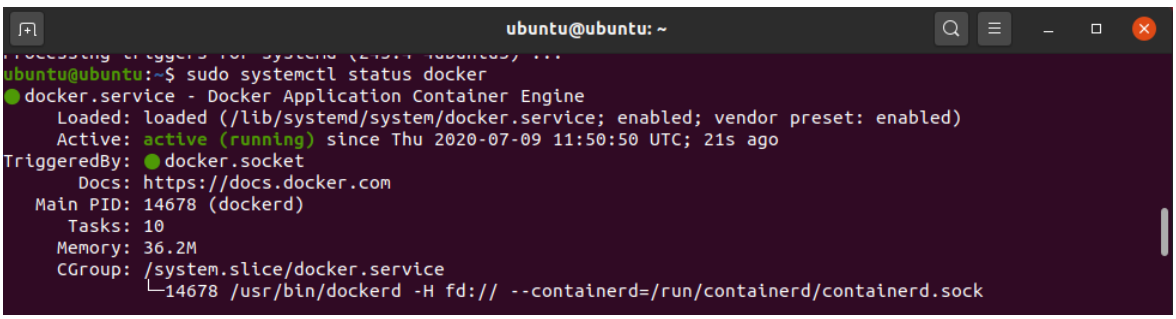


Fig. 4. Docker status.

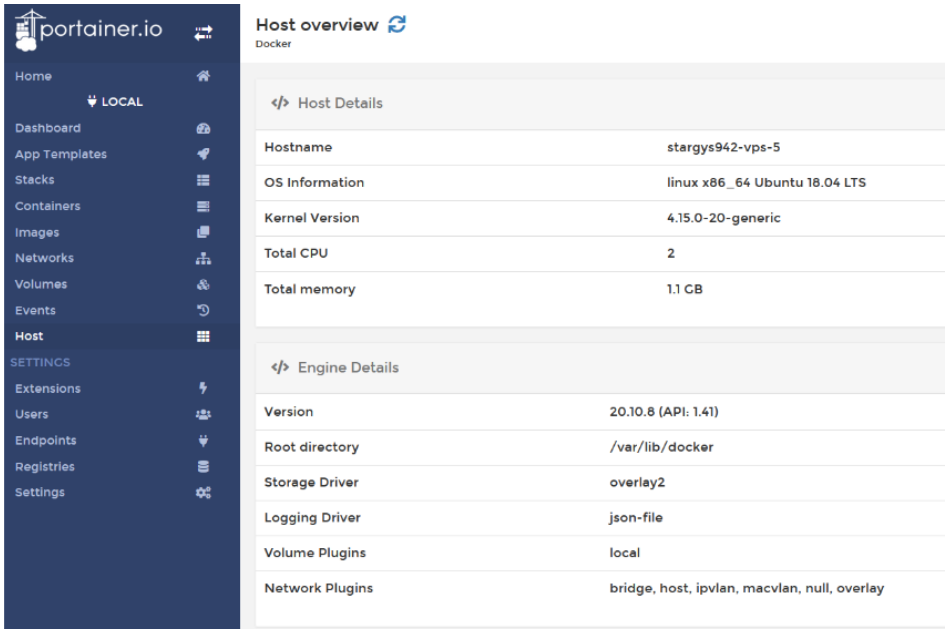


Fig. 5. Running a docker application.

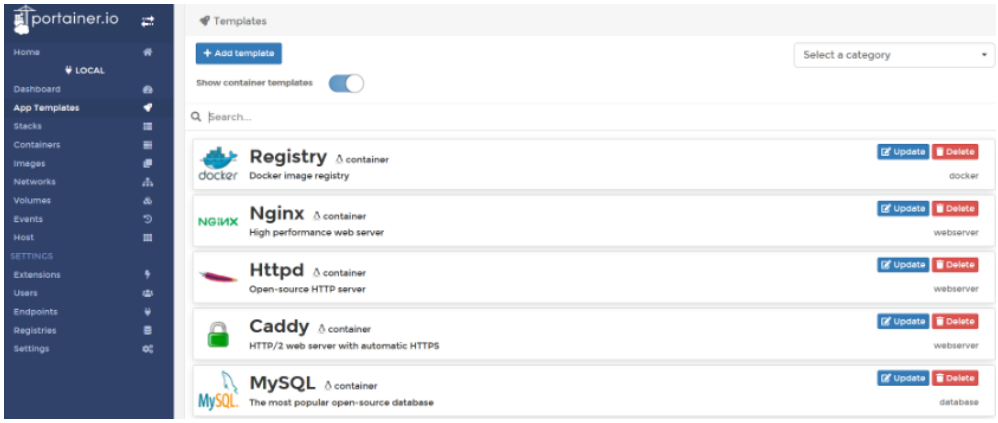


Fig. 6. App template.

In the scenario when we use standard TLS, traffic can be available for analysis and modification; we can see this by switching to the "Decrypted SSL Data" tab. (Fig. 7).

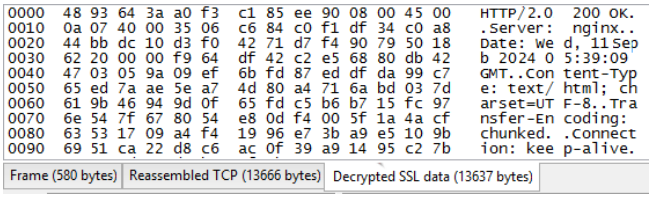


Fig. 7. Decrypted SSL data.

After setting up modified TLS on the VoIP server, the traffic is encrypted and cannot be analyzed or spoofed (Fig. 8).

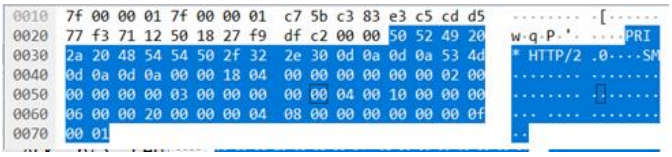


Fig. 8. Ercrypted SSL data.

IV. Discussion

The integration of lattice-based cryptography into the TLS protocol provides a significant enhancement to the overall security of the IP PBX Asterisk system. Lattice-based algorithms such as Learning With Errors (LWE) offer resistance to quantum computing attacks, which is critical as traditional cryptographic methods like RSA and ECC are vulnerable to quantum threats. By using post-quantum key exchange algorithms, the system ensures that even if a quantum computer were able to break traditional cryptographic schemes, it would not be able to compromise the confidentiality of the communications.

Moreover, the combination of AES for symmetric encryption and HMAC for message integrity ensures a strong layer of defense against eavesdropping and tampering in real-time VoIP communications. This safeguards the system from Man-in-the-Middle (MITM) attacks, where an attacker might attempt to intercept or alter voice and data transmissions. The proposed system's use of lattice-based cryptographic keys during the TLS handshake further secures this process by

making key exchanges immune to both classical and quantum-based cracking attempts.

Additionally, the system's architecture, which incorporates Docker for isolated and modular environments, enhances security by reducing the attack surface. Each module operates within a container, minimizing the risk of one compromised module affecting others. The containerized environment also ensures that vulnerabilities in one part of the system do not compromise the whole network, adding a layer of isolation and protection.

However, the system's security also depends on proper management of cryptographic keys and certificates, as failure to securely store or manage these can still leave the system vulnerable to traditional attacks. While lattice-based cryptography provides robust protection against future quantum threats, it is essential to ensure that the deployment and management of the cryptographic infrastructure are secure and well-maintained to prevent other potential security breaches, such as phishing or insider threats.

The proposed system offers several advantages, including quantum-resistant security through lattice-based cryptography, ensuring long-term protection against future quantum attacks. It enhances key exchange protocols using the LWE algorithm, secures real-time communication via post-quantum TLS, and maintains data integrity through HMAC. The system's modular architecture allows for easy integration and scalability. However, it also has some disadvantages, such as increased computational overhead, higher latency in real-time communications, and potential challenges with integration into existing infrastructure. The large key sizes of lattice-based cryptography may also impact storage and bandwidth efficiency. Furthermore, the lack of widespread standardization in post-quantum cryptography adds complexity to its implementation and maintenance.

V. Conclusion

This research makes a significant contribution to the development of cybersecurity systems for IP networks by proposing a practical implementation of post-quantum lattice-based cryptography in the TLS protocol. The main contribution of the research is the integration of lattice cryptographic algorithms into the Asterisk IP PBX infrastructure, which provides reliable protection against threats associated with the development of quantum computers. Unlike existing approaches that focus on the theoretical aspects of post-quantum cryptography, this study demonstrates, the practical application of these methods in real-world conditions using virtual machines and Docker for performance and security testing.

It also contributes to the development and evaluation of a system capable of protecting real-time data transmission with minimal delays, which is especially important for IP telephony and VoIP systems. The implementation of multi-layered data protection using a modified TLS protocol not only increases the security of transmitted information, but also reduces the likelihood of successful man-in-the-middle (MITM) attacks, which is confirmed by experimental data. Furthermore, the work opens up new possibilities for further research and implementation of post-quantum methods in critical communication systems.

Future research in the field of cybersecurity for IP networks and IP PBXs can focus on several directions. First, post-quantum cryptographic algorithms such as lattice systems should be further optimized to reduce their computational complexity and improve their efficiency in real-world conditions, especially for high-load IP telephony. Second, it is worth exploring the possibilities of integrating other post-quantum cryptography methods (based on isogenies or multivariate polynomials) into existing infrastructures to improve security.

Acknowledgment

This article was prepared as part of the government contract as requested by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan on the subject formulated as «Development of a method for improving the security of a telecommunications network based on IP-PBX Asterisk» (project No. AP14871745).

References

[1] V. Mavroeidis, K.Vishi, M.D. Zych, and A, Jøsang, "The Impact of Quantum Computing on Present Cryptography," *Int. J. of Adv. Comput. Sci. and Appl. (IJACSA)*, vol. 9, no.3, 2018. Doi: 10.14569/Ijacsa.2018.090354.

[2] Ch. Majdoubi, S.El Mendili. and Youssef Gahi, "Quantum Cryptology in the Big Data Security Era." *Int. J. of Adv. Comput. Sci. and Appl. (IJACSA)*, vol. 15, no. 7, 2024. Doi:10.14569/Ijacsa.2024.0150761.

[3] Y. Baseri, V. Chouhan, and A.Hafid, "Navigating quantum security risks in networked environments: A comprehensive study of quantum-safe network protocols," *Comput. & Secur.*,vol. 142, 2024, 103883. Doi: 10.1016/j.cose.2024.103883.

[4] A. Zunussov, A. Baikenov, O. Manankova, T. Zheltaev, and T. Zhaksylyk, "Quality of service management in telecommunication network using machine learning technique," *Indonesian J. of Electr. Eng. and Comput. Sci.*, vol. 32, no. 2, pp. 1022–1030, 2023. Doi: 10.11591/ijeecs.v32.i2.pp1022-1030.

[5] S.B. Hegde, A.Jamuar, and R. Kulkarni, "Post Quantum Implications on Private and Public Key Cryptography," In Proceedings of the 2023 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES), Tumakuru, India, 7–8 July 2023; pp. 1–6.

[6] S. V. Singh., D.Kumar, "Enhancing Cyber Security Using Quantum Computing and Artificial Intelligence: A Review," *Int. J. of Adv. Res. in Sc.,* Communication and Technology, pp. 4-11. Doi: 10.48175/ijarsct-18902.

[7] M.Kumar, "Post-quantum cryptography Algorithm's standardization and performance analysis," Array, vol. 15, 2022, 100242, Doi: 10.1016/j.array.2022.100242.

[8] J. Hekkala, M. Muurman, and K. Halunen, "Implementing Post-quantum Cryptography for Developers," *Sn Comput. Sci.*, vol. 4, no. 365, 2023. Doi:10.1007/s42979-023-01724-1.

[9] A. Horpenyuk, I. Opirskyy, and P. Vorobets, "Analysis of Problems and Prospects of Implementation of Post-Quantum Cryptographic Algorithms," *CEUR Workshop Proceedings*, pp.39-49. 2023.

[10] T. Tripathi, A. Awasthi, Sh. Pratap Singh, and A. Chaturvedi, "Post Quantum Cryptography & its Comparison with Classical Cryptography," arXiv:2403.19299v1, 2024.

[11] S. Ricci, P. Dobias, L. Malina, J. Hajny and P. Jedlicka, "Hybrid Keys in Practice: Combining Classical, Quantum and Post-Quantum Cryptography," *in IEEE Access*, vol. 12, pp. 23206-23219, 2024, doi: 10.1109/ACCESS.2024.3364520.

[12] N. Aviram, B. Dowling, I. Komargodski, K. G. Paterson, E. Ronen and E. Yogev, "Practical (post-quantum) key combiners from one-wayness and applications to TLS," *Cryptol. ePrint Arch.*, pp. 1-24, Feb. 2022.

[13] M. R. Albrecht, B. R. Curtis, A. Deo, A. Davidson, R. Player, E. W. Postlethwaite, et al., "Estimate all the LWE NTRU schemes!", Proc. Secur. Cryptogr. Netw. 11th Int. Conf. (SCN), pp. 351-367, Sep. 2018.

[14] V. Dinesh Reddy, P. Ravi, Ashu Abdul, Mahesh Kumar Morampudi and Sriramulu Bojjagani, "Techniques for Solving Shortest Vector Problem" *Int. J. of Adv. Comput. Sci. and Appl. (IJACSA)*, vol.12, no. 5, 2021. Doi: 10.14569/Ijacsa.2021.0120598.

[15] A. Langley, "Real-world measurements of structured-lattices and supersingular isogenies in TLS", 2019

[16] R. Overbeck, and N. Sendrier, "Code-based cryptography," In: Bernstein, D.J., Buchmann, J., Dahmen, E. (eds) Post-Quantum Cryptography. Springer, Berlin, Heidelberg, 2009. Doi: 10.1007/978-3-540-88702-7_4.

[17] D. J. Bernstein, T. Lange, C. Peters, and H. C. A. van Tilborg, "Explicit bounds for generic decoding algorithms for code-based cryptography," In A. Kholosha, E. Rosnes, and M. Parker, editors, Pre-proceedings of WCC 2009, pages 168–180, Bergen, 2009.

[18] K. Preetha Mathew, S. Vasant, and C. Pandu Rangan, "ON PROVABLY SECURE CODE-BASED SIGNATURE AND SIGNCRYPTION SCHEME," *IACR Cryptology ePrint Archive*, 2012:585, 2012.

[19] T. Matsumoto and H. Imai, "Public Quadratic Polynomial-Tuples for Efficient Signature Verification and Message Encryption," *Adv. in Cryptology*, Springer, pp. 419-453, 1988.

[20] J. Ding and A. Petzoldt, "Current State of Multivariate Cryptography," *IEEE Security & Privacy*, vol. 15, no. 4, pp. 28-36, 2017. Doi: 10.1109/MSP.2017.3151328.

[21] C. Tao et al., "Simple Matrix Scheme for Encryption", *Post-Quantum Cryptography* (PQCrypto 13), pp. 231-242, 2013.

[22] A. K, D. S. de Oliveira, J. L´opez, and R, Cabral, "High Performance of Hash-based Signature Schemes" *Int. J. of Adv. Comput. Sci. and Appl. (IJACSA)*, vol. 8, no. 3, 2017. Doi: 10.14569/Ijacsa.2017.080358.

[23] K. Bicakci, K. Ulker, Y.Uzunay, H. T. Şahin, and M. S. Gündoğan, "Quantum-Resistance Meets White-Box Cryptography: How to Implement Hash-Based Signatures against White-Box Attackers?," *IACR Communicat. in Cryptology*, vol. 1, no. 2, Jul 08, 2024. Doi: 10.62056/an59qgxq.

[24] D. A. Cooper, D. C. Apon, Q. H. Dang, M. S. Davidson, M. J. Dworkin, and C. A. Miller, "Recommendation for stateful hash-based signature schemes," *NIST Special Publication*, 800:208, 2020. Doi: 10.6028/NIST.SP.800-208.

[25] L. Li, X. Lu, and K. Wang, "Hash-based signature revisited," *Cybersecurity*, vol. 5, no. 1, pp. 1–26, 2022. Doi: 10.1186/s42400-022-00117-w.

[26] J. Buchmann, E. Dahmen, and M. Szydlo, "Hash-based Digital Signature Schemes. In: Bernstein, D.J., Buchmann, J., Dahmen, E. (eds) Post-Quantum Cryptography. Springer, Berlin, Heidelberg. doi: 10.1007/978-3-540-88702-7_3.

[27] L. De Feo, "Mathematics of Isogeny Based Cryptography," arXiv:1711.04062, 2017. Doi: 10.48550/arXiv.1711.04062.

[28] M. Campagna et al., "Supersingular isogeny key encapsulation," ed, 2019.

[29] F.Tellez, and J. Ortiz, "Comparing AI Algorithms for Optimizing Elliptic Curve Cryptography Parameters in e-Commerce Integrations: A Pre-Quantum Analysis" *Int. J. of Adv. Comput. Sci. and Appl. (IJACSA)*, vol. 15, no. 6, 2024. Doi: 10.14569/Ijacsa.2024.01506153.

[30] V.Rao M. and S. Malladi, "Secure Energy Efficient Attack Resilient Routing Technique for Zone based Wireless Sensor Network" *Int. J. of Adv. Comput. Sci. and Appl. (IJACSA)*, vol. 11, no. 12, 2020. Doi: 10.14569/Ijacsa.2020.0111267.

[31] H. Abroshan, "A Hybrid Encryption Solution to Improve Cloud Computing Security using Symmetric and Asymmetric Cryptography Algorithms" *Int. J. of Adv. Comput. Sci. and Appl. (IJACSA)*, vol. 12, no. 6, 2021. Doi: 10.14569/IJACSA.2021.0120604.

[32] C. V. Manjushree, and A. N. Nandakumar, "A Hybrid Double Encryption Approach for Enhanced Cloud Data Security in Post-Quantum Cryptography," *Int. J. of Adv. Comput. Sci. and Appl. (IJACSA)*, vol. 14, no. 12, 2023. Doi: 10.14569/IJACSA.2023.0141225.

[33] G. Omar, A. and Sh. Kamal, "A Survey on Cryptography Algorithms," *Int. J.of Sci. and Res. Public.* (IJSRP), 2018. Doi:10.29322/IJSRP.8.7.2018.P7978.

[34] M. Yakubova, O. Manankova, A. Mukasheva, A. Baikenov, and T. Serikov, "The Development of a Secure Internet Protocol (IP) Network Based on Asterisk Private Branch Exchange (PBX)," *Appl. Sci.* (Switzerland), vol. 13, no. 19, 2023. Doi: 10.3390/app131910712.

[35] O.A. Manankova, M.Z. Yakubova, M.A. Rakhmatullaev, and A.S.Baikenov, "Simulation of the Rainbow Attack on the SHA-256 Hash function," *J. of Theoret. and Appl. Inf. Tech.*, vol. 101, no. 4, pp. 1594–1603, 2023.

[36] M.Yakubova, T. Serikov, O. Manankova, "Development and Research of a Method for MultiLevel Protection of Transmitted Information in IP Networks Based on Asterisk IP PBX Using Various Codecs," *Int. J. of Adv. Comput. Sci. and Appl. (IJACSA)*, vol. 15, no. 7, pp. 724–731, 2024. Doi: 10.14569/IJACSA.2024.0150771.

[37] M. A. Al-Shabi, "A Survey on Symmetric and Asymmetric Cryptography Algorithms in information Security," *Int. J.of Sci. and Res. Public.*, vol. 9, no. 3, pp.576-589, 2019. Doi: 10.29322/IJSRP.9.03.2019.p8779.

[38] V. Ganeshkar, and M.Kulkarni, "QUANTUM CRYPTOGRAPHY FOR A SECURE COMMUNICATION," *Int. J. of Res. In Comput. Appl. and Inf. Techn.*, vol. 7, no. 1, January-June 2024, pp. 17-29, Article ID: IJRCAIT_07_01_003.

[39] F. Opiłka, M. Niemiec, M. Gagliardi, and M. A. Kourtis, "Performance Analysis of Post-Quantum Cryptography Algorithms for Digital Signature," *Appl. Sci.*, vol.14, no. 12, 2024; Doi: 10.3390/app14124994.

[40] P. Shrivastava, K.K. Soni, and A. Rasool, "Evolution of Quantum Computing Based on Grover's Search Algorithm," In Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, pp. 1–6, 6–8 July 2019.

Energy Optimization Management Scheme for Manufacturing Systems Based on BMAPPO: A Deep Reinforcement Learning Approach

Zhe Shao*
Woosong University, Endicott College, Korea, 34606

***Abstract*—To address the depletion of traditional energy sources and the increasingly severe environmental pollution, countries around the world have accelerated the deployment of renewable energy generation equipment. Energy optimization management for microgrids can address the randomness of factors such as renewable energy generation and load, ensuring the safe and stable operation of the system while achieving objectives such as cost minimization. Therefore, this paper conducts an in-depth study of energy optimization management schemes for microgrids and designs a multi-microgrid energy optimization management model and algorithm based on deep reinforcement learning. For the joint optimization problem among multiple microgrids with power flow between them, a two-layer energy optimization management scheme based on the multi-agent proximal policy optimization (PPO) algorithm and optimal power flow (BMAPPO) is proposed. This scheme is divided into two layers: first, the lower layer uses the multi-agent proximal policy optimization algorithm to determine the output of various controllable power devices in each microgrid; then, based on the lower layer's optimization results, the upper layer uses a second-order cone relaxation optimal power flow model to solve the optimal power flow between multiple microgrids, achieving power scheduling among them; finally, the total cost of the upper and lower layers is calculated to update the network parameters. Experimental results show that compared with other schemes, the proposed scheme achieves multi-microgrid energy optimization management at the lowest cost while ensuring online execution speed.**

Keywords*—Microgrid; energy optimization management; deep reinforcement learning; multi-agent; Proximal Policy Optimization (PPO)

I. INTRODUCTION

Electric power is an indispensable driving force in modern society. In recent years, with the rapid development of technology and the continuous growth of the global population, the demand for electricity has been increasing year by year [1]. To meet this demand, the current smart grid is transitioning towards a more structured system based on microgrids. This transition, which optimizes energy storage systems through collaboration and self-organization, is key to driving the existing energy system towards being more intelligent, robust, and green.

Microgrid-based smart grids are not only better at integrating emerging distributed components but also position microgrids as an effective part of distribution and transmission system management through the evolving flexibility markets and new grid management concepts. The flexibility provided by microgrids, as well as their ability to operate in both grid-connected and island modes, are crucial solutions to the challenges faced by future transmission and distribution networks. The introduction of microgrids helps improve the reliability of power systems, reduce emissions, and expand the energy options for future power systems. Furthermore, the multi-microgrid structure formed by the interconnection of microgrids enhances the resilience, security, and intelligence of energy systems, supporting energy systems that incorporate large amounts of variable renewable energy.

Although the technologies related to components such as renewable energy generation and storage systems in microgrids have matured, joint optimization in microgrids remains challenging due to uncertainties in renewable energy generation, load, and energy prices. In addition to dealing with the fluctuations in renewable energy and load, storage devices must be optimized and controlled according to their operating costs or physical constraints. When multiple microgrids need to be optimized simultaneously, the complexity of the algorithms also increases. Given the technical and economic advantages of microgrids in future energy systems, ensuring the efficient and stable operation of microgrids has become a hot research topic.

A multi-microgrid refers to a system where multiple individual microgrids within a certain area are interconnected to achieve power mutual assistance. Compared to a single microgrid, a multi-microgrid has several advantages: First, a multi-microgrid can integrate large-scale renewable energy generation equipment, achieving a higher penetration rate of renewable energy through power flow between microgrids; second, it allows for the shared use of devices such as energy storage and generation, enabling a microgrid with large-capacity energy storage or high-power generation equipment to supply energy to other microgrids when the main grid's electricity price is high, further reducing costs; third, it enhances system robustness, allowing energy to be sourced from other microgrids if a microgrid's supply equipment fails or if the main grid experiences a power outage. Therefore, from the perspective of economic efficiency and future development trends, it is necessary to conduct research on energy optimization management for multi-microgrids.

Microgrid energy optimization management is an important research area in the power industry, aiming to achieve intelligent control and autonomous scheduling decisions for microgrids through optimization techniques. Under the premise

of ensuring the safe operation of equipment, the output of controllable power devices in microgrids is optimized to cope with fluctuations in renewable energy generation, load, and real-time electricity prices, meet load demand, avoid power waste, and minimize operational costs.

For the scenario of multiple microgrids, this paper proposes an energy optimization management scheme based on multi-agent proximal policy optimization. Since a single microgrid has limited capacity to deal with system uncertainties and often needs to trade with the main grid, joint optimization of multiple microgrids is expected to become a future development trend. To this end, this paper proposes a dual-layer structure for multi-microgrid energy optimization management. In the lower layer, the output of devices in each microgrid is decided based on a multi-agent proximal policy optimization algorithm, where centralized training ensures optimization effectiveness and decentralized decision-making protects user privacy. In the upper layer, the optimal power flow between microgrids is solved using a second-order cone relaxation optimal power flow model, ensuring power mutual assistance between microgrids.

The organization of this paper is as follows: Section II provides a detailed review of existing microgrid energy optimization management schemes; Section III proposes a dual-layer energy optimization management scheme based on multi-agent reinforcement learning to further improve the stability of microgrid operation and reduce costs; Section IV verifies the feasibility of the proposed method through case studies; Section V summarizes the contributions of the entire paper.

II. Literature Review

This section will review existing work related to resource management to highlight the gaps in current research.

A. Related Works

Energy optimization management in microgrids is essentially a constrained optimization problem with uncertain factors. Currently, the methods used in the field of microgrid energy optimization management mainly include metaheuristic algorithms (e.g., genetic algorithms), mathematical programming methods (e.g., mixed-integer linear programming), robust optimization, stochastic optimization, model predictive control, and deep reinforcement learning algorithms.

Metaheuristic algorithms have been widely applied in the field of microgrid energy optimization management. Among them, genetic algorithms and particle swarm optimization are the most commonly used, with similar techniques including ant colony optimization [2], crow search algorithm [3], and simulated annealing algorithm [4], among others. Torkan et al. [5] applied a multi-objective genetic algorithm to the optimization management of microgrids, considering the uncertainties brought by demand response (DR) programs, reactive power loads, and renewable energy. They optimized microgrid operations with energy cost and greenhouse gas emissions as objective functions, while this optimization objective function was constrained by a series of system constraints and was solved using a genetic algorithm.

In mathematical programming-based schemes, mixed-integer linear programming (MILP) can handle optimization problems where variables are continuous and discrete, making it highly suitable for application in microgrid energy optimization management. MILP can be used to establish mathematical models of microgrid components and optimize the cost function. Sigalo et al. [6] proposed an energy management scheme for grid-connected microgrids focused on battery storage systems, considering changes in grid electricity prices, renewable energy generation, and load demand, and determined the charging and discharging power of the battery to minimize the overall energy loss cost.

Stochastic and robust optimization-based microgrid energy optimization management schemes have been proposed to address the stochastic factors and prediction errors inherently present in microgrids. Chen et al. [7] proposed a new cumulative regret-based robust optimization method for the optimal management of grid-connected multi-energy microgrids considering uncertainty factors. Compared to traditional robust optimization methods, the proposed strategy ensures the robustness of microgrids and reduces the conservatism of microgrid operations. Additionally, by considering the demand response of thermal loads, the optimization model for microgrid energy management was improved. Abunima et al. [8] proposed a two-stage microgrid optimization scheduling method that coordinates microgrid assets under uncertainty, allowing microgrid operators to save operational costs without increasing investment costs, while meeting load demand. Nair et al. [9] considered an islanded microgrid composed of photovoltaic generation, supercapacitors, and regenerative fuel cells, utilizing a model predictive control algorithm. The goal was to enhance the utilization of renewable energy, improve microgrid operational efficiency, and reduce the degradation rate of the storage system.

A common feature of the above model-based methods is their reliance on precise predictions of uncertain factors in microgrids. Once prediction errors occur, the performance of these methods can be significantly impacted. Additionally, the computational cost of these methods is typically high, facing the issue of "curse of dimensionality"; as the complexity of the optimization problem increases, the computational cost multiplies, making it difficult to meet the real-time requirements of microgrid energy optimization management. To address these issues, some researchers have employed deep reinforcement learning (DRL) to solve the problem of microgrid energy optimization management. Deep reinforcement learning is a data-driven or model-free algorithm that does not rely on precise modeling of the microgrid environment. Thanks to the powerful perception capabilities of deep learning algorithms, DRL can effectively learn the microgrid environment model. Additionally, due to the strong decision-making ability of reinforcement learning, it can efficiently solve optimization problems. Alabdullah et al. [10] proposed a microgrid energy management solution based on the Deep Q-Network (DQN) algorithm, considering the stochastic behavior of various factors in the microgrid and modeling different grid components, while adhering to various power flow constraints in real-world environments.

B. Research Gaps and Motivation

Based on the above literature review, the following conclusions can be drawn:

1) Real-time performance and stability challenges: Current microgrid energy optimization methods, including mathematical programming, stochastic optimization, robust optimization, and MPC, struggle with high computational complexity and real-time performance due to the handling of uncertainties. MPC, in particular, faces difficulties in ensuring stability and has not fully accounted for model uncertainties.

2) Dependency on accurate predictions: Model-based methods rely heavily on accurate predictions of uncertainties in microgrids. Prediction errors and high computational costs can lead to performance issues and the "curse of dimensionality," making real-time energy optimization challenging.

3) Need for further research in deep reinforcement learning: While deep reinforcement learning offers potential advantages such as real-time scheduling and a general framework, its applicability across different microgrid architectures and its effectiveness in online optimization require further research and validation.

Addressing these issues is critical for improving the effectiveness and applicability of microgrid energy management, which is the focus of this paper.

III. Multi-Microgrid Dual-Layer Energy Optimization Management Model Based on Reinforcement Learning

A. Multi-Microgrid Dual-Layer Energy Optimization Management Model

The structure of the multi-microgrid dual-layer energy optimization management model designed in this chapter is shown in Fig. 1. The lower layer consists of N microgrids, each containing photovoltaic generation equipment, wind power generation equipment, energy storage devices, micro gas turbines, loads, and a control center. The microgrids are interconnected through energy routers.

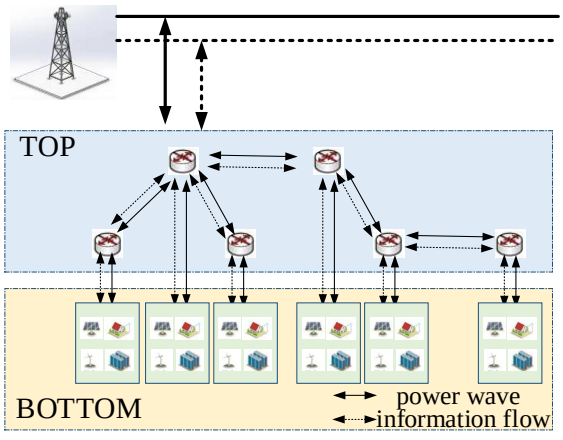


Fig. 1. Schematic diagram of the multi-microgrid dual-layer structure.

The upper layer is an abstracted topology based on the lower layer, which facilitates power flow analysis. In the process of one round of multi-microgrid energy optimization management, the lower layer first uses multi-agent deep reinforcement learning to decide the output of controllable power devices [11]. Then, the upper layer calculates the optimal power flow based on the regulated results of the lower layer. Finally, based on the results of the two-layer optimization, the reward value is calculated, and the neural network parameters are updated. Therefore, the following sections will first introduce the multi-agent deep reinforcement learning algorithm MAPPO in the lower layer, then explain the optimal power flow model for the multi-microgrid in the upper layer, and finally present the overall algorithm flow and experimental results.

B. Lower Layer Multi-Agent Deep Reinforcement Learning Algorithm

As shown in Fig. 2, MAPPO is a multi-agent variant of PPO (Proximal Policy Optimization) and operates using a Centralized Training with Decentralized Execution (CTDE) approach. In centralized training, the Critic network of each agent can use global information during the offline training phase to achieve better convergence. In decentralized execution, each agent's Actor network can only observe its own state to make decisions during the online execution phase. Extensive experiments have shown that the clipped form of PPO consistently outperforms the penalized form, and hence, the clipped form is adopted in this work. The optimization objective can be written as follows:

$$\max \hat{\mathbb{E}}_t\left[\min\left(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta),1-\varepsilon,1+\varepsilon)\hat{A}_t\right)\right] \quad (1)$$

$$\text{clip}(r_t(\theta),1-\varepsilon,1+\varepsilon) = \begin{cases} 1-\varepsilon, & r_t(\theta) < 1-\varepsilon \\ 1+\varepsilon, & r_t(\theta) > 1+\varepsilon \\ r_t(\theta), & \text{other} \end{cases} \quad (2)$$

Here, clip(·) is the clipping function. When $\hat{A}_t > 0$, it indicates that the action a_t taken at this moment is better than the average, so maximizing Eq. (1) will increase $r_t(\theta)$, meaning the probability of action a_t in the new policy will increase. However, $r_t(\theta)$ will not increase beyond 1+ε. Conversely, when $\hat{A}_t < 0$, it indicates that the action a_t taken at this moment is worse than the average, so maximizing Eq. (1) will decrease $r_t(\theta)$, meaning the probability of action a_t in the new policy will decrease. However, $r_t(\theta)$ will not decrease below 1-ε.

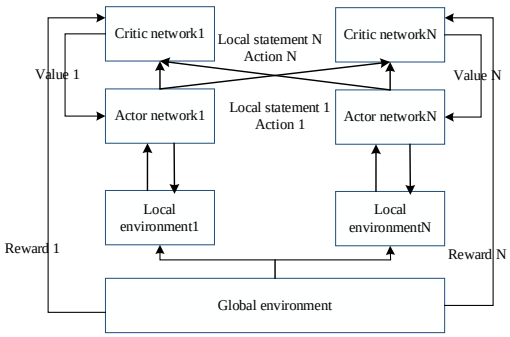


Fig. 2. MAPPO algorithm architecture.

C. Upper-Layer Optimal Power Flow Model

Currently, distribution networks are primarily radial in structure. A radial structure with multiple microgrids can be modeled using the Branch Flow Model (BFM). Fig. 3 illustrates a schematic diagram of the Branch Flow Model. For node j:

- V_j represents the voltage at the node;
- $s_j = p_j + iq_j$represents the power injection at the node.

For the branch from node i to node j (i→j):

- I_{ij} represents the branch current;
- $S_{ij} = P_{ij} + iQ_{ij}$ represents the power at the sending end of the branch;
- $Z_{ij} = r_{ij} + ix_{ij}$ represents the branch impedance.

The topology of the upper-level multi-microgrid system is denoted as G(N,E)G(N,E). Finally, by applying angle relaxation and second-order cone relaxation, the optimal power flow problem is transformed into a convex optimization problem. Specifically:

$$l_{ij} \geq \frac{P_{ij}^2 + Q_{ij}^2}{v_i}, \forall (i,j) \in E \Leftrightarrow \left\| \begin{bmatrix} 2P_{ij} \\ 2Q_{ij} \\ l_{ij} - v_i \end{bmatrix} \right\|_2 \leq l_{ij} + v_i, \quad \forall (i,j) \in E \tag{3}$$

At this point, the optimization variables for the optimal power flow problem become $\{p_i, q_i, P_{ij}, Q_{ij}, l_{ij}, v_i\}$. For a radial network, the literature has proven that angle relaxation is tight; under the conditions that the objective function is strictly increasing and convex, the second-order cone relaxation is also tight. At this stage, the optimal power flow problem has been modeled as a convex optimization problem, which can be conveniently solved using commercial solvers like Gurobi.

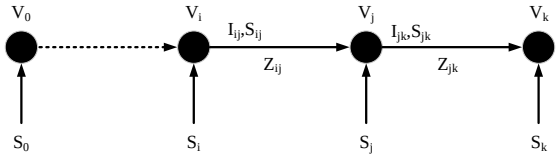


Fig. 3. Schematic diagram of branch flow structure.

D. BMAPPO: Multi-Microgrid Dual-Layer Energy Optimization Management Algorithm

The constraints in a multi-microgrid system arise from three aspects: the electrical boundary constraints of various devices in the microgrid, the power balance constraints of each microgrid, and the power flow constraints between microgrids [12]. For convenience, unless otherwise stated, i represents the ith microgrid in the multi-microgrid system; t represents the tth round of optimization management; N represents the set of microgrid nodes; E represents the set of edges (power lines) between microgrids.

1) Device boundary constraint

a) Micro gas turbine boundary constraints:

$$P_{i,\min}^{MT} < P_{i,t}^{MT} < P_{i,\max}^{MT}, \forall i \in N \tag{4}$$

In the above formula, $P_{i,t}^{MT}$ is the output power of the micro gas turbine (kW); $P_{i,\min}^{MT}$ and $P_{i,\max}^{MT}$ are the lower and upper bounds of the micro gas turbine's output power (kW).

b) Energy storage device boundary constraints:

$$P_{i,\min}^{ESS} < P_{i,t}^{ESS} < P_{i,\max}^{ESS}, \forall i \in N \tag{5}$$

$$SOC_{i,\min} \leq SOC_{i,t} \leq SOC_{i,\max} \ \forall i \in N \tag{6}$$

In the above formulas, $P_{i,tS}^{ESS}$ is the charging/discharging power of the energy storage device (kW); $P_{i,\min}^{\mathrm{ESS}}$ and $P_{i,\max}^{ESS}$ are the lower and upper bounds of the charging/discharging power of the energy storage device (kW); $SOC_{i,t}$ is the state of charge of the energy storage device; $SOC_{i,\min}$ and $SOC_{i,\max}$ are the lower and upper bounds of the state of charge of the energy storage device.

c) Main grid boundary constraints

$$P_{i,\min}^{MG} < P_{i,t}^{MG} < P_{i,\max}^{MG}, \forall i \in N \tag{7}$$

$$Q_{i,\min}^{MG} < Q_{i,t}^{MG} < Q_{i,\max}^{MG}, \forall i \in N \tag{8}$$

In the above formulas, $P_{i,t}^{MG}$ is the active power traded between the microgrid and the main grid (kW); $P_{i,\min}^{MG}$ and $P_{i,\max}^{MG}$ are the lower and upper bounds of the active power traded between the microgrid and the main grid (kW); $Q_{i,t}^{MG}$ is the reactive power traded between the microgrid and the main grid (kVar); $Q_{i,\min}^{MG}$ and $Q_{i,\max}^{MG}$ are the lower and upper bounds of the reactive power traded between the microgrid and the main grid (kVar).

2) Power balance constraints

a) Active power balance constraint

$$p_{i,t} = P_{i,t}^{PV} + P_{i,t}^{WT} + P_{i,t}^{MT} + P_{i,t}^{ESS} + P_{i,t}^{MG} - P_{i,t}^{Load}, \forall i \in N \tag{9}$$

In the above formula, $p_{i,t}$ is the injected active power (kW); $P_{i,t}^{PV}$ and $P_{i,t}^{WT}$ are the active power outputs of photovoltaic and wind power generation, respectively (kW); $P_{i,t}^{MT}$ is the output power of the micro gas turbine (kW); $P_{i,t}^{ESS}$ is the charging/discharging power of the energy storage device (kW); $P_{i,t}^{MG}$ is the active power traded between the microgrid and the main grid (kW); $P_{i,t}^{\mathrm{Load}}$ is the active power of the load (kW).

b) Reactive power balance constraint

$$q_{i,t} = Q_{i,t}^{MG} - Q_{i,t}^{Lood}, \forall i \in N \tag{10}$$

In the above formula, $q_{i,t}$ is the injected reactive power (kVar); $Q_{i,t}^{\mathrm{MG}}$ is the reactive power from the main grid (kVar); $Q_{i,t}^{\mathrm{Load}}$ is the reactive power of the load (kVar).

3) Power flow constraints

$$v_j = v_i - 2\left(r_{ij}P_{ij} + x_{ij}Q_{ij}\right) + \left(r_{ij}^2 + x_{ij}^2\right)l_{ij}, \forall (i,j) \in E \tag{11}$$

$$p_j = \sum_{k:j\to k} P_{jk} - \sum_{i:i\to j} \left(P_{ij} - r_{ij}l_{ij}\right), \forall j \in N \tag{12}$$

$$q_j = \sum_{k:j\to k} Q_{jk} - \sum_{i:i\to j} \left(Q_{ij} - x_{ij}l_{ij}\right), \forall j \in N \tag{13}$$

$$l_{ij} \geq \frac{P_{ij}^2+Q_{ij}^2}{v_i},\ \forall(i,j) \in E \quad (14)$$

$$\left|I_{ij}\right| \leq \bar{I_{ij}},\ \forall(i,j) \in E \quad (15)$$

$$\underline{V_i} \leq |V_i| \leq \bar{V_i},\ \forall i \in N \quad (16)$$

$$\underline{s_i} \leq s_i \leq \bar{s_i},\ \forall i \in N \quad (17)$$

E. Optimization Objective Function

In this chapter, considering the uncertainties in renewable energy generation, load, and electricity prices, the objective is to minimize the cooperative operation cost of multiple microgrids. An energy optimization management model for multiple microgrids is constructed, and the objective function is as follows:

$$\min F_t = \min \sum_{t=1}^{T} \left(\sum_{i=1}^{N} \left(F_{i,t}^{MG} + F_{i,t}^{MT} + F_{i,t}^{ESS}\right) + F_t^{\text{Loss}}\right) \quad (18)$$

Where:

- F_t is the total operating cost of the multi-microgrid system.
- t represents the t-th round of optimization management.
- T is the total number of rounds within an energy optimization management period.
- $F_{i,t}^{MG}$ represents the cost of trading with the main grid.
- $F_{i,t}^{MT}$ represents the generation cost of the micro gas turbine.
- $F_{i,t}^{ESS}$ represents the loss cost of the energy storage device.
- F_t^{Loss} represents the power transmission loss between microgrids.

For simplicity, in this chapter, ΔtΔt represents the time length of one round of optimization management (hours).

a) Cost of trading between microgrid and main grid

$$F_{i,t}^{MG} = c_t^{MG} P_{i,t}^{MG} \cdot \Delta t \quad (19)$$

In the above formula, ctMGctMG is the electricity price of the main grid during the tt-th round of optimization management ($/kWh), and P(i,t)MGP(i,t)MG is the active power purchased by the microgrid from the main grid (kW).

b) Micro gas turbine generation cost

$$F_{i,t}^{MT} = \left(a \cdot \left(P_{i,t}^{MT}\right)^2 + b \cdot P_{i,t}^{MT} + c\right) \cdot \Delta t \quad (20)$$

In the above formula, aa, bb, and cc are cost coefficients.

c) Energy storage device loss cost

$$F_{i,t}^{ESS} = \left(c^{\text{ESS}} \cdot \left(P_{i,t}^{cha} \cdot \eta_{\text{cha}} + P_{i,t}^{dis} / \eta_{dis}\right)\right) \cdot \Delta t \quad (21)$$

In the above formula, cESScESS is the loss coefficient; P(i,t)chaP(i,t)cha and P(i,t)disP(i,t)dis are the charging and discharging powers of the energy storage device, respectively (kW); ηchaηcha and ηdisηdis are the charging and discharging efficiencies of the energy storage device, respectively.

d) Transmission loss between microgrids

$$F_t^{\text{Loss}} = \sum_{ij \in E} c_t^{MG} \left|I_{ij,t}\right|^2 r_{ij} \cdot \Delta t \quad (22)$$

In the above formula, $I_{ij_{,,}}$ and r_{ij} are the current (A) and impedance (Ω) between microgrid i and microgrid j, respectively.

Based on the above description, solving the optimization problem in Eq. (19) requires handling optimization variables that can be divided into lower-layer and upper-layer optimization variables. The lower-layer optimization variables include:

- The generation power of micro gas turbines in microgrids $\{P_{i,t}^{MT}\}$.
- The charging and discharging power of energy storage devices in microgrids $\{P_{i,t}^{ESS}\}$.
- The active power traded between the microgrid and the main grid $\{P_{i,t}^{\text{MG}}\}$.
- The reactive power traded between the microgrid and the main grid $\{Q_{i,t}^{\text{MG}}\}$.

The upper-layer optimization variables include:

- Node voltage $\{v_{j,t}\}$.
- Node injected power $\{p_{j,t}, q_{j,t}\}$.
- Branch current $\{l_{ij,t}\}$ and branch power $\{P_{ij,t}, Q_{ij,t}\}$.

F. Construction of a Partially Observable Markov Decision Process

The multi-agent reinforcement learning algorithm MAPPO (Multi-Agent Proximal Policy Optimization) is used to solve the multi-microgrid energy optimization management problem, which can be modeled as a Partially Observable Markov Decision Process (POMDP). POMDP can be defined as a five-tuple $(\mathcal{S}, \{\mathcal{O}_i\}_{i=1}^n, \{\mathcal{A}_i\}_{i=1}^n, \mathcal{P}, \{r_i\}_{i=1}^n)$, where:

- $\mathcal{S}$ is the global state space.
- $\mathcal{P}$ is the state transition function.
- For agent i, the observation space is $\mathcal{O}_i$, the action space is $\mathcal{A}_i$, and the reward function is r_i.

Further, in multi-agent reinforcement learning, the interaction process between agents and the environment is as follows:

- At time step t, each agent i obtains an observation state $o_{i,t} \in \mathcal{O}_i$ and selects an action $a_{i,t} \in \mathcal{A}_i$ according to the policy $\pi_i: \mathcal{O}_i \times \mathcal{A}_i \rightarrow [0,1]$.
- Then, the system transitions to the next state $o_{i,t+1}$ according to the state transition probability $\mathcal{P}$ and receives a reward $(s_t, a_{i,t})$.

The objective of multi-agent reinforcement learning is to maximize the cumulative return: $J(\pi) = \mathbb{E}\left[\sum_{t=0}^{T} \frac{1}{n} \gamma^t \sum_{i=1}^{n} r_{i,t}\right]$.

Therefore, the multi-microgrid energy optimization management problem is modeled as a POMDP below.

a) Observation space definition

$$o_{i,t} = \left[P_{i,t}^{PV}, P_{i,t}^{WT}, P_{i,t}^{Load}, c_t^{MG}, SOC_{i,t}\right] \quad (23)$$

In the above formula:

- $P_{i,t}^{PV}$ and $P_{i,t}^{WT}$ are the photovoltaic and wind power generation outputs, respectively.
- $P_{i,t}^{Load}$ is the load power.
- c_t^{MG} is the real-time electricity price of the main grid.
- $SOC_{i,t}$ is the state of charge of the energy storage device.

b) b) State Space Definition

$$s_t = \left[o_{1,t}, o_{2,t}\ \mathrm{K}, o_{n,t}\right] \quad (24)$$

The state space contains the observation space of all agents and represents the global information of the multi-agent environment.

c) Action space definition

$$a_{i,t} = \left[P_{i,t}^{MT}, P_{i,t}^{ESS}, P_{i,t}^{\mathrm{MG}}, Q_{i,t}^{\mathrm{MG}}\right] \quad (25)$$

In the above formula:

- $P_{i,t}^{MT}$ is the power output of the micro gas turbine.
- $P_{i,t}^{ESS}$ is the charging/discharging power of the energy storage device.
- $P_{i,t}^{\mathrm{MG}}$ is the active power traded between the microgrid and the main grid.
- $Q_{i,t}^{\mathrm{MG}}$ is the reactive power traded between the microgrid and the main grid.

d) Reward function definition

$$r_{i,t} = -\left(\alpha \cdot \left(F_{i,t}^{MG} + F_{i,t}^{MT} + F_{i,t}^{ESS}\right) + \beta \cdot \sum_{i=1}^{N} \mathrm{cons}\left(SOC_{i,t}\right)\right) \quad (26)$$

$$r_t = -F_t^{\mathrm{Loss}} + \sum_{i=1}^{N} r_{i,t} \quad (27)$$

Eq. (26) is the reward function for agent ii, where α and β are coefficients, and cons (·) is a penalty function introduced in the previous chapter. Eq. (27) is the total reward for the multi-microgrid system. Since the algorithm designed in this chapter is a dual-layer optimization, the total reward is calculated after the upper-layer optimization is completed. It includes the sum of all lower-layer agent rewards and the upper-layer optimization objective F_t^{Loss}. It is worth mentioning that Eq. (23) can also be decomposed using the vectorization approach proposed in the previous chapter, but for simplicity, the traditional reward function is used here. Online Decision-Making Process of the MAPPO-Based Multi-Microgrid Dual-Layer Energy Optimization Management Scheme. The online decision-making process is shown in Algorithm 1.

Algorithm 1: Multi-Microgrid Energy Optimization Management Algorithm Online Decision-Making Process

Input: Actor network weights θπθπ

Output: Values of all optimization variables in the tt-th round of multi-microgrid energy optimization

Step 1: Obtain the initial state of all agents.

For t=1,2,…,T do:

Step 1: Each agent i∈N in the lower layer observes state o_t^i and selects an action $a_{i,t}$ based on the policy, obtaining the optimization variables $\left\{P_{i,t}^{MT}, P_{i,t}^{ESS}, P_{i,t}^{\mathrm{MG}}, Q_{i,t}^{\mathrm{MG}}\right\}, i \in N$.

Step 2: The upper layer solves the optimal power flow problem using a commercial optimizer, obtaining the optimization variables $\left\{v_{j,t}\right\} \setminus \left\{p_{j,t}, q_{j,t}\right\} \setminus \left\{\Delta P_{i,t}^{MG}, \Delta Q_{i,t}^{MG}\right\}, i \in N$ and $\left\{l_{ij,t}\right\} \setminus \left\{P_{ij,t}, Q_{ij,t}\right\}, ij \in E$.

Step 3: Apply all optimization variables to the multi-microgrid system and obtain the next state s_{t+1}.

End for

IV. Experimental Results Analysis

A. Simulation Environment and Parameter Settings

The IEEE 33-bus system structure, as shown in Fig. 4, is used as the upper-level topology for the dual-layer energy optimization management of multiple microgrids. Microgrids can be placed at any node, and the impedance value between two adjacent nodes is the average of the branch impedances between the nodes. For example, if microgrids are connected at nodes 1 and 21 to form a system with two microgrids, the impedance between nodes 1 and 21 would be the average impedance of branches 1-18, 18-19, 19-20, and 20-21. This setup allows the generation of various simulation environments on the IEEE 33-bus system. In this chapter, nodes 1, 2, 5, and 24 are selected to connect microgrids 1 to 4, forming a multi-microgrid structure with four microgrids. Energy optimization management is performed every hour.

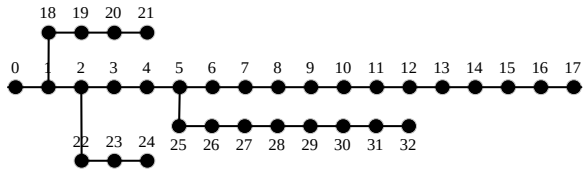


Fig. 4. Schematic diagram of the IEEE 33-bus system structure.

The experimental equipment used in this chapter includes an Intel(R) Core(TM) i5-10210U CPU @1.60GHz 2.10 GHz and an NVIDIA GeForce RTX2060. The compiler used is Pycharm 2022.3, and the programming language is Python 3.8. The commercial optimizer used for solving the upper-level optimal power flow is the Python version of Gurobi 10.0.1. The implementation framework for the lower-level multi-agent deep reinforcement learning algorithm is the mainstream neural network development framework Pytorch. The MAPPO parameters are set as shown in Table I:

TABLE I. MAPPO PARAMETER SETTINGS

Parameter	Value
Discount Factor γ	0.95
Number of Neurons in Hidden Layer	128
Actor Network Learning Rate l_a	0.0003
Critic Network Learning Rate l_c	0.0001
Clipping Function Hyperparameter ε	0.2
Size of Experience Replay Pool D	10000
Mini-batch Size B	96
Maximum Number of Training Epochs	20000

B. Optimization Management Results Analysis

1) Convergence analysis: As shown in Fig. 5, during the offline training process, the BMAPPO algorithm proposed in this chapter converges at around 9000 iterations. The smoothed cumulative reward oscillates around -2250. Since the reward function is a negative function of the optimization management cost, the convergence curve shows an upward trend, indicating that during the iteration process, the agent learns effective strategies to reduce the cost of multi-microgrid optimization management. During offline training, 80% of the dataset was used. Next, the network parameters after training are saved, and the remaining 20% of the dataset is used to simulate the online optimization management process to test the effectiveness of the algorithm.

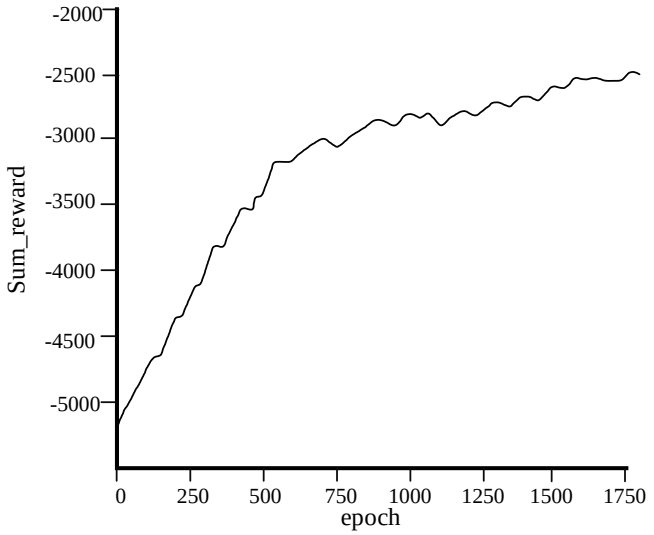


Fig. 5. Convergence curve of the BMAPPO algorithm during offline training.

2) Effectiveness analysis: Fig. 6 shows the lower-level optimization results for Microgrid 1 and Microgrid 2.

Fig. 6(a1) and 6(a2) represent the output curves of controllable power devices in Microgrid 1 and Microgrid 2, respectively, including the power generation of micro gas turbines, the charging and discharging power of energy storage devices, and the power regulation results of transactions with the main grid. Combined with the real-time electricity price fluctuations of the main grid shown in Fig. 5, it can be seen that when the main grid electricity price is low from 1 to 5 hours, the power of the energy storage devices in both microgrids is negative (except for a discharge of about 40 kW in Microgrid 1 at 5 hours), representing charging of the energy storage devices. During this period, a large amount of electricity is purchased from the main grid, while the micro gas turbines almost do not output power. When the main grid price is moderately high from 9 to 17 hours, the trend of output from controllable power devices is similar to that from 1 to 5 hours, but the load power gap in this period is also relatively low, so the power purchased from the main grid is also relatively less. When the electricity price is high from 6 to 8 hours and 18 to 22 hours, the output of the energy storage devices is generally positive, representing discharging of the energy storage devices. During these periods, the power purchased from the main grid is at a low point, and the micro gas turbines output a large amount of power to fill the load power gap. Therefore, the BMAPPO proposed in this chapter is effective in cost savings.

Fig. 6(b1) and 6(b2) show the state of charge (SoC) curves of the energy storage devices in Microgrid 1 and Microgrid 2, respectively, with the upper and lower bounds of the SoC indicated, set at 0.9 and 0.1. An SoC less than 0.1 indicates an over-discharging condition, while an SoC greater than 0.9 indicates an over-charging condition. It can be seen that the energy storage device in Microgrid 2 is in a safe state during the typical day, with no over-charging or over-discharging occurring. However, Microgrid 1 shows a slight over-charging condition from 16 to 19 hours.

Fig. 7 illustrates the upper-level optimization results for Microgrid 1 and Microgrid 2.

Fig. 7(a1) and 7(a2) display the power injection curves for Microgrid 1 and Microgrid 2, respectively. Positive power injection indicates that there is excess power within the microgrid that is not being consumed, while negative power injection indicates that there is an unsatisfied power deficit within the microgrid. The upper-level power flow optimization will balance the power injection between microgrids, ensuring power balance in each microgrid. From Fig. 7(a1), it can be seen that Microgrid 1 has positive power injection from 17 to 21 hours, which can be transmitted to other microgrids. From Fig. 7(a2), it can be seen that Microgrid 2 has negative power injection from 18 to 22 hours, which can be obtained from other microgrids. This shows that the upper-level microgrids can achieve mutual support.

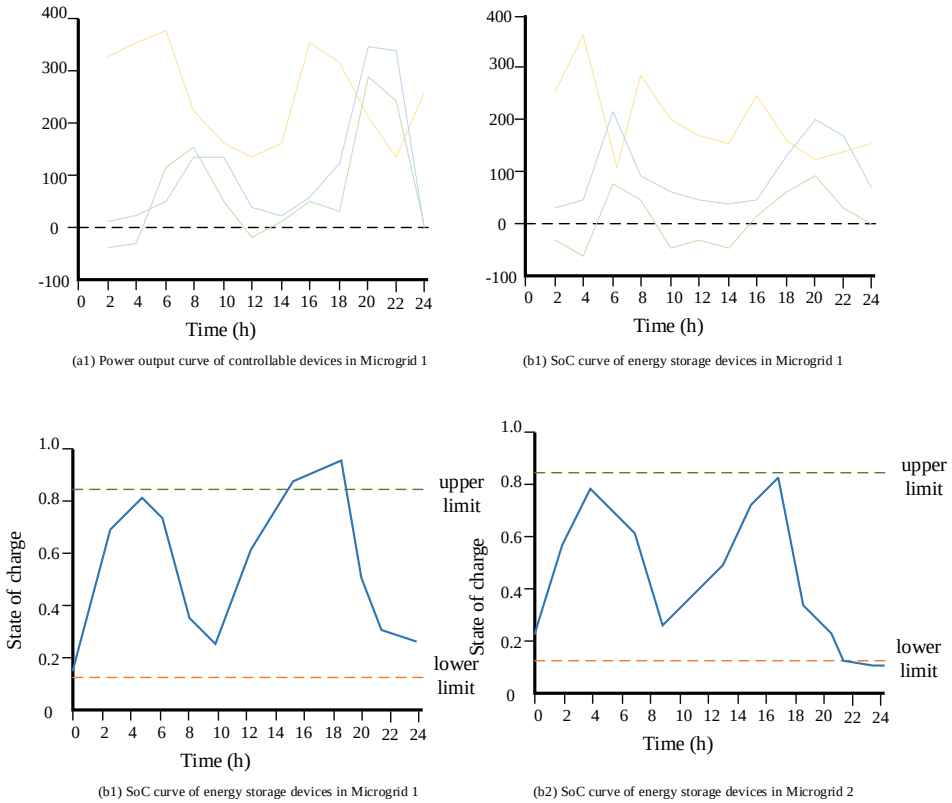


Fig. 6. Lower-level optimization results of Microgrid 1 and Microgrid 2.

Fig. 7. Upper-level optimization results of Microgrid 1 and Microgrid 2.

V. CONCLUSION

This paper proposes a dual-layer optimization management scheme based on the multi-agent reinforcement learning algorithm MAPPO for the energy optimization management problem of multiple microgrids. The lower layer uses MAPPO to make decisions on the power output of each microgrid device, handling power imbalances, while the upper layer achieves overall power balance of multiple microgrids through a second-order cone relaxation optimal power flow model. The experimental results show that the designed BMAPPO algorithm effectively achieves mutual support between microgrids and significantly reduces the energy optimization management costs of multiple microgrids.

Although the deep reinforcement learning-based energy optimization management scheme proposed in this paper shows significant advantages in cost savings, there is still room for improvement:

The deep reinforcement learning method relies on a large amount of historical data for training, which may be difficult to obtain in practical applications. Therefore, future research should focus on reducing dependence on historical data or improving data utilization efficiency.

This paper only studies the situation where microgrids are connected to the main grid, while islanded microgrids have widespread applications in remote areas, where their optimization management cost control is of significant importance. Therefore, how to reduce the management costs of islanded microgrids while ensuring safety is an important direction for future research.

ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g.” Avoid the stilted expression, “One of us (R. B. G.) thanks . . .” Instead, try “R. B. G. thanks.”

REFERENCES

[1] Fu T, Liu S, Li P. Intelligent smelting process, management system: Efficient and intelligent management strategy by incorporating large language model[J]. Frontiers of Engineering Management, 2024: 1-17.

[2] Liu S, Zheng P, Shang S. A novel bionic decision-making mechanism for digital twin-based manufacturing system[J]. Manufacturing Letters, 2023, 35: 127-131.

[3] Liu S, Zheng P. A Novel Bionic Digital Twin-Based Manufacturing System Toward the Mass Customization Paradigm[C]//2023 IEEE 19th International Conference on Automation Science and Engineering (CASE). IEEE, 2023: 1-6.

[4] ANGELIM J, AFFONSO C. Energy management on university campus with photovoltaic generation and BESS using simulated annealing[C]. proceedings of the 2018 IEEE Texas Power and Energy Conference (TPEC). IEEE, 2018: 1-6.

[5] TORKAN R, ILINCA A, GHORBANZADEH M. A genetic algorithm optimization approach for smart energy management of microgrids [J]. Renewable Energy, 2022, 197: 852-63.

[6] SIGALO M B, PILLAI A C, DAS S, et al. An energy management system for the control of battery storage in a grid-connected microgrid using mixed integer linear programming [J]. Energies, 2021, 14(19): 6212.

[7] CHEN T, CAO Y, QING X, et al. Multi-energy microgrid robust energy management with a novel decision-making strategy [J]. Energy, 2022, 239: 121840.

[8] ABUNIMA H, PARK W-H, GLICK M B, et al. Two-Stage stochastic optimization for operating a Renewable-Based Microgrid [J]. Applied Energy, 2022, 325: 119848.

[9] NAIR U R, COSTA-CASTELLó R. A model predictive control-based energy management scheme for hybrid storage system in islanded microgrids [J]. IEEE access, 2020, 8: 97809-22.

[10] ALABDULLAH M H, ABIDO M A. Microgrid energy management using deep Q-network reinforcement learning [J]. Alexandria Engineering Journal, 2022, 61(11): 9069-78.

[11] Fu T, Li P, Liu S. An imbalanced small sample slab defect recognition method based on image generation[J]. Journal of Manufacturing Processes, 2024, 118: 376-388.

[12] Fu T, Liu S, Li P. Digital twin-driven smelting process management method for converter steelmaking[J]. Journal of Intelligent Manufacturing, 2024: 1-17.

Design Science Research: Applying Integrated Fogg Persuasive Frameworks to Validate Rural ICT Design Requirements

Design for Digital Equality in Rural Areas

Noela Jemutai Kipyegen[1], Benard Okelo[2]
School of Informatics and Innovative Systems, Jaramogi Oginga Odinga University of Science and Technology, Kisumu, Kenya[1]
School of Biological, Physical, Mathematics and Actuarial Sciences,
Jaramogi Oginga Odinga University of Science and Technology, Kisumu, Kenya[2]

***Abstract*—Designing for digital equality is critical in the modern world. Digital inequality is more pronounced in rural areas, where the majority are illiterate and poor. As a result of this, individuals are not motivated, enabled and triggered enough to access and use Information and Communication Technologies. Additionally, existing rural ICT artifacts or applications are not usable by these demographics. Therefore, this paper understood and validated rural ICT design requirements. It achieved this by developing a community learning system, applying design science research methodology and integrated Fogg persuasive frameworks. Results shows that use of local language, local content, videos, audio, touch-based input, proper content categorization, accessibility (location) and peer participation and collaboration fosters user engagement with the ICT artifact. These approaches had a significant impact in the achievement of user self-efficacy. This is explained by the task findings, 71%, 83%, 70% and 78% successfully, within the stipulated time and on their own, accomplished tasks 1, 2,3 and 4 respectively. Users found the content to be practical and applicable to their day-to-day activities. Users appreciated the system's potential for learning indicating that it could significantly enhance their knowledge and skills. The significance of Design Science Research and integrated Fogg persuasive frameworks in creating usable and accessible ICT solutions tailored to the needs of the target population cannot be underrated. It was concluded that design solutions targeting vulnerable demographics are key to the success of designs for digital equality. In other words, usable solutions for the aged, women illiterate, uneducated and the poor, are more usable for the young, men, literate, educated, and the rich (financially stable). Thus, enhancing inclusivity in access and use of rural ICTs.**

Keywords*—Digital equality; rural areas; design requirements; rural ICT; artifacts; validate; Fogg frameworks; design science research

I. INTRODUCTION

Digital inequalities have emerged as a growing concern in modern societies [1]. The disparities pertain to differences in access and use of ICTs such as internet connectivity, internet-enabled devices and digital literacy skills among individuals living in rural areas [2, 3]. According to study [2], these three factors are essential for communities to create a sustainable access to the digital world, specifically as critical aspects of society such as education, workforce development and innovation transition to online platforms [2]. Therefore, the absence of robust, and all-encompassing applications or information access makes it difficult for people to participate in the digital community and access digital services, which remains exclusive on a systemic level. Thus, there is a need for targeted ICT solutions which will enhance inclusivity in access and use among various demographics in the community.

Even if people have equal access to digital technologies, such as the presence of rural ICT centers, disparities in skills, motivation, and literacy levels can result in digital inequality. Consequently, marginalized groups (such as the illiterate and uneducated) of all ages tend to avoid adopting modern technologies, leading them to miss out on important information. Encouraging these groups, who have little to no exposure to technology, to integrate ICTs into their daily lives is a significant challenge. This underscores the necessity for a thoughtful design approach. According to researcher [Rogers], technology can be utilized to influence people's behavior. Hence, we applied Fogg frameworks to validate ICT design requirements for rural areas and to identify additional requirements unique to rural areas.

The term persuasive technology pertains to interactive information technology that is specifically created to change users' attitudes or behavior or both within the realm of human-computer interaction [4]. Given its manipulative nature and ability to exert influence in a subtle manner [5], particularly on individuals who may lack literacy or motivation, it has the potential to significantly impact the design of digital solutions for rural areas. This helps to achieve equality in access and use of ICTs.

This research is part of a larger study aiming to establish a framework for creating rural ICTs that promote inclusivity. This paper specifically focuses on validating the requirements to achieve this broader objective. It starts by providing the study's background and related work in before delving into the requirements for designing for illiterate and semi-illiterate individuals, as the primary aim is to validate these requirements. The validation process enhances the comprehension of the design requirements for rural areas, ultimately leading to the development of a robust, practical, and applicable framework. Requirements were sourced from existing research,

incorporated into the Fogg persuasive framework, and used to design persuasive features in a community learning system. The conclusion discusses the persuasive features identified through evaluation that promote inclusivity in accessing and using rural ICTs. Lastly, future work is recommended.

II. Background and Related Work

A. Fogg Persuasive Frameworks

According to study [6], users perceive computers as tools, media and social actors. This is known as captology (computer as a persuasive technology) triad.

The Fogg Behavioral Model (FBM) is a conceptual framework used to understand human behavior. This framework is particularly important for those involved in studying and creating persuasive technology. Persuasive technology focuses on automating behavior change. To effectively incorporate experiences that promote behavior change, designers need a thorough understanding of human psychology, especially the factors that drive human behavior. The FBM provides designers and researchers with a systematic way to consider the underlying factors of behavior change. It highlights that motivation, ability, and triggers must all align at the same time for the behavior to occur; otherwise, it will not happen. The inclusion of FBM in the UX design process assists designers in creating user-centered designs and encourages meaningful and impactful interaction with the product [7].

Fogg's persuasive integrated framework merges the triad elements of captology with the components of FBM. Captology focuses on the characteristics of ICT, while FBM emphasizes human factors. Fig. 1 demonstrates, using computer as a tool, how these components will be integrated to create a robust framework that will assist in defining the design requirements for rural ICT. MF is motivation features, AF is ability features and TF is triggers features, all to the computer as a tool.

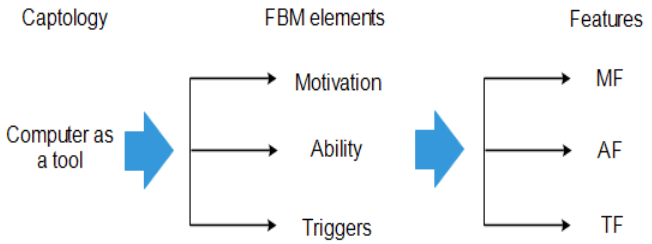


Fig. 1. Illustration of integrated Fogg persuasive frameworks.

The Chomoka Mobile Application, developed by [8], is a relevant persuasive technology for this study. It was designed specifically for Village Savings and Loan Associations (VSLAs) in seven rural African countries. Their goal in utilizing FBM was to build a more user-friendly and intuitive application, focusing on creating triggers that promote consistent usage. Furthermore, this approach improved the overall user experience and impact of the Chomoka app. It is evident that incorporating the Fogg Behavior Modeling framework could have further enhanced the success of the Chomoka app.

In their content analysis study, [9] systematically assessing the prevalence of persuasive design elements in popular early childhood apps [9] found out that a majority of them displayed a high prevalence of passive motivation features such as vibrant colors and music, as well as ability features like in-game help suggestions and repetitive tasks. In their conclusion, the researchers emphasized the importance of considering the developmental consequences of persuasive design features in apps and investigating how they might influence problematic screen behaviors in early childhood.

In their study, [10], demonstrated how the mobile app Sedentaware can prevent sedentary behavior by encouraging users to engage in physical activities. They illustrated how their model can complement a mathematical model to develop more effective persuasive technologies. Their research highlighted a key drawback of persuasive technology theory: it lacks the necessary level of detail for designers to systematically create persuasive technologies. This led us to combine Fogg frameworks with DSRM.

Researchers, [11] combined Fogg Behavioral Model and the Hook model to design features for an app as a component of a larger persuasive system to help improve three key areas of study habits: study scheduling, class preparation and group study. The proposed mobile companion app formed the overall persuasive system which was used to influence student study habits to improve their learning outcomes.

B. Design Science Research Methodology

Researcher, [12] conducted a study with the goal of reducing the digital divide between literate and illiterate populations in Iraq. They achieved this by creating a mobile app focused on educating about traffic signs. The app was developed using the Design Science Research (DSR) method. Upon development and evaluation, it was found that the illiterate participants demonstrated highly competent performance in completing tasks. The study confirmed that the app, which incorporated design principles such as voice instructions, symbols, and the local language, was effective, efficient, and garnered high user satisfaction. This study demonstrates that developing mobile apps for illiterate individuals can enhance their technological knowledge and bridge the gap between them and those with access to modern technology.

In their study, [13] created a mobile app to help illiterate individuals in Bangladesh find jobs, aiming to bridge the digital gap in developing nations. They used the DSRM to develop and assess the IT tool. They discovered that incorporating minimal text and voice prompts was crucial for designing mobile apps for illiterate users, and that using voice instead of text for information input was another essential design principle for this demographic. They also emphasized the importance of using images and culturally relevant symbols in User Interfaces (UI) to improve user understanding of the app's functions. In their future work, they suggested conducting a study to address the issues identified in their research, such as using methods like observation and on-site visits in addition to semi-structured interviews for requirement gathering, and considering a longitudinal study for broader insights. They also highlighted the need to recruit a larger number of participants for future studies.

C. Existing Gap

Numerous research efforts have aimed to narrow the digital divide for people who are illiterate by concentrating on the design of user interfaces and interaction methods for providing digital services in different global settings. Nonetheless, it is clear that no prior research in Kenya has investigated and validated these requirements for creating a practical design solution, which contributes to digital inequality among different population groups in rural areas. Furthermore, our study combines a persuasive framework with the DSRM process, demonstrating a theoretical aspect that is not commonly found in other studies that strive to promote equality through designing for individuals with limited literacy skills, particularly, rural areas.

D. ICT Design Requirements for Rural Areas

Illiteracy is a significant challenge in rural areas [14]. Moreover, as outlined by study [15], the increase in computer illiteracy among individuals is significantly contributing to the expansion of the digital divide. According to [16], literacy is defined as "the capacity to securely and effectively access, manage, comprehend, integrate, communicate, evaluate, and generate information using digital technologies. It could be for the purpose of gaining knowledge, employment, securing decent jobs, and entrepreneurship." This literacy encompasses a continuous process of learning and mastery in reading, writing, and numerical skills throughout one's lifespan, forming part of a broader range of competencies that encompass digital literacy, media literacy, education for sustainable development, global citizenship, and specific skills required for various professions [16].

Despite the educational deficiencies prevalent among a substantial portion of the population, numerous existing applications and their interfaces cater primarily to individuals with higher levels of education. These interfaces tend to be text based thus making them unsuitable for individuals who are illiterate, semi-literate, or lack digital literacy, thereby limiting their ability to effectively engage with software systems. Consequently, these sizable demographics are unable to derive the advantages offered by ICT services if they are unable to access them, ultimately giving rise to digital inequality, the divergence between individuals who possess unrestricted access to information and communication technologies and those who either lack access entirely or have limited access to such technologies. Hence, the task of bridging the digital gap would be unattainable in the absence of implementing ICTs for this extensive population of disadvantaged (illiterate and semi-literate) cohorts [17].

Researchers, [18] revealed various aspects that need to be recognized when designing persuasive systems include responsiveness, error-freeness, ease of access, ease of use, convenience, information quality, positive user experience, attractiveness, user loyalty, and simplicity, to name a few; however, they indicate that more precise requirements for software qualities will have to be defined to be able to communicate the ideas from idea generators and/or management to software engineers. Additionally, [18] suggested four principles of persuasive design: (1) primary task (applying reduction, tunneling, tailoring, personalization, self-monitoring, simulation, and rehearsal). (2) dialogue (include, praise, rewards, reminders, suggestion, similarity, liking, and social role). (3) system credibility (include, trustworthiness, expertise, surface credibility, real-world feel, authority, third-party endorsements, and verifiability). (4) social support (social facilitation, social comparison, normative influence, social learning, cooperation, competition, and recognition).

On the other hand, [6] indicated that changing peoples' attitudes and behavior require;

- tailored information.
- computer simulations.
- application of animate characteristics (like, physical features, emotions, voice communication),
- play animate roles (like, coach, pet, assistant, opponent), or
- social rules or dynamics (like greetings, apologies, taking turns)

Numerous studies were undertaken to pioneer innovative ICT-based solutions aimed at delivering information to illiterate and semi-illiterate residents of rural areas in developing countries [12, 15, 17, 18, 19]. They indicated the following requirements;

- Applying minimized hierarchical navigation
- Uses of text-free UI in local language;
- speech/voice and touch based interactions
- apply intuitive icons, symbols, images, and photographs
- use audio, video, and animated instruction/guide;
- include context sensitive suggestions
- proactive briefing
- local content (useful application)

E. Validating Rural ICT Design Requirements

Validating the rural ICT design needs, entail designing an intervention (artifact) for solving the problem [20]. There are various design methodoogies and frameworks that can be applied in the development of ICTs for inclusivity. [21], revealed that appropritate design techniques for ICT inventions encourage participartory design. Utilizing a participatory action research approach in ICT design and development is essential for not only preserving local culture and identity but also fostering the local creation of ICT [22]. This design has been found to be appropriate for the development of innovations for rural areas and especially systems where usability is a critical factor [23, 24, 25].

Persuasive design employs participatory design (PD) method, which involves potential users as active participants in the design process for creating computer systems and computer-based activities [26]. In addition to providing the venerable with the opportunity to design ICT products and services, PD fosters a positive relationship between users and designers, allowing designers to leverage the users' creativity and knowledge about

the context of use, thereby fostering a sense of ownership of the resulting ICT product. Additionally, it is a design technique which helps in creating successful human-computer-interactions where designers are required to play a big role as agents of influence [6].

DSR is another approach that involves creating a new practical object to tackle a broad category of issues and then assessing its effectiveness in solving problems within that category [27]. It utilizes a participatory approach. Researchers, [12, 13] applied DSRM in the design and validation of design requirements for illiterate and semi-illiterate users.

III. METHODOLOGY

This study employed Design Science Research methodology (DSRM). The selection of DSRM was based on its rigor, flexibility and cost effectiveness [28]. DSRM has six steps; problem identification, objectives for the solution, design and development, demonstration, evaluation, and communication [29]. This methodology is systemic and flexible, with a focus on improvement through iterative processes that involve empathetic observation, understanding user needs, and devising innovative approaches to meet those needs. The objective is to transition from explaining phenomena to creating interventions to solve problems. Therefore, design science is especially valuable for identifying and implementing practical solutions within complex situations [20] like the case of rural areas.

As a result of DSR's flexibility and support for participatory design, this study was able to integrate persuasive theory into its process. The urban informatics design model [30] was adapted, see Fig. 2. Their model was designed specifically for a socio-cultural context, hence shared similarities with rural informatics.

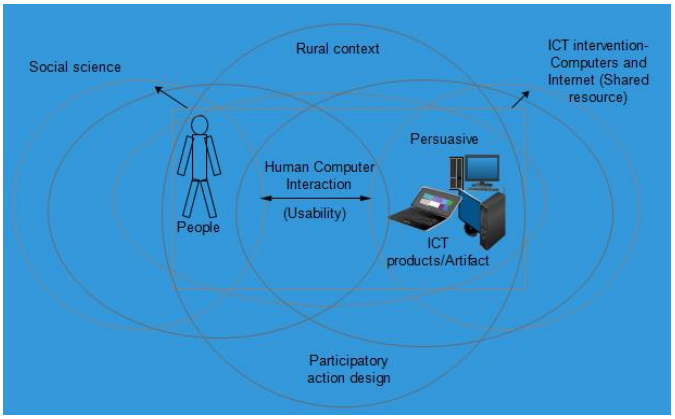


Fig. 2. Adapted PADR for rural Informatics.

For ease of integrating participatory action design research (PADR) with DSR, we simplified it and the resultant model is shown in Fig. 3.

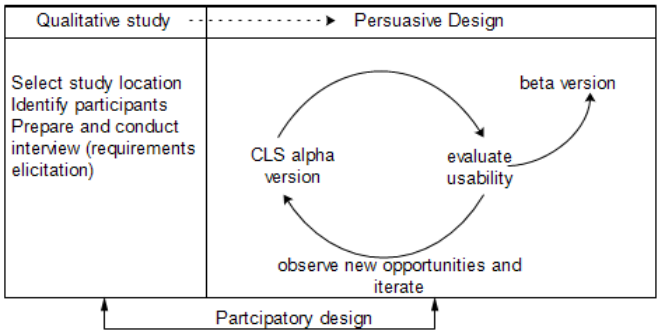


Fig. 3. Rural ICT design and development model.

A. Persuasive Design

Studies [33, 34] have revealed that as a result of affordability and illiteracy issues, individuals in rural areas are not persuaded into accessing and utilizing ICTs. Hence, the need for persuasive design methodology. Furthermore, persuasive design approaches have been applied by other information systems researchers [8, 9, 11, 18, 35, 36, 37].

The use of persuasive systems can involve either persuading humans through computers or using computers to mediate persuasion. It is true that the idea of a persuader is quite intricate because computers themselves do not have intentions; it is the people who develop, disseminate, or use the technology that have the intention to influence someone's attitudes or actions. [38]. Additionally, [38], stated that Fogg's framework and principles offer valuable ways to comprehend persuasive technology. According to these statements, we applied a persuasive framework that leverages both Fogg Behavior Model (FBM) and his Captology (Computer as a persuasive technology) Triad framework. This was also based on the understanding that designing Information and Communication Technology solutions for illiterate users requires a thoughtful approach. The FBM emphasizes Motivation (desire to participate in access and use of RICTs), Ability (the ease of accessing and using RICTs), and Triggers or prompts (cues that prompt users to access and use RICTs). On the other hand, the functional captology triad outlines how computers can act as tools (enhances users' abilities to perform tasks), mediums (conveys or simulates experiences), and social actors (engages users in social interaction). As seen in Table VI the integrated Fogg persuasive frameworks incorporate the design requirements for illiterate found in the existing literature.

B. Application of Design Science Research

1) Problem-centered approach: The study is problem centered which was necessitated by the need to identify and validate rural ICT design requirements. On the quest of how to design rural ICT solutions for inclusivity, this study obtained good information but barely none was from Kenyan setup, it was mostly based in other countries like Asian and developed nations. Fig. 4 illustrates this approach.

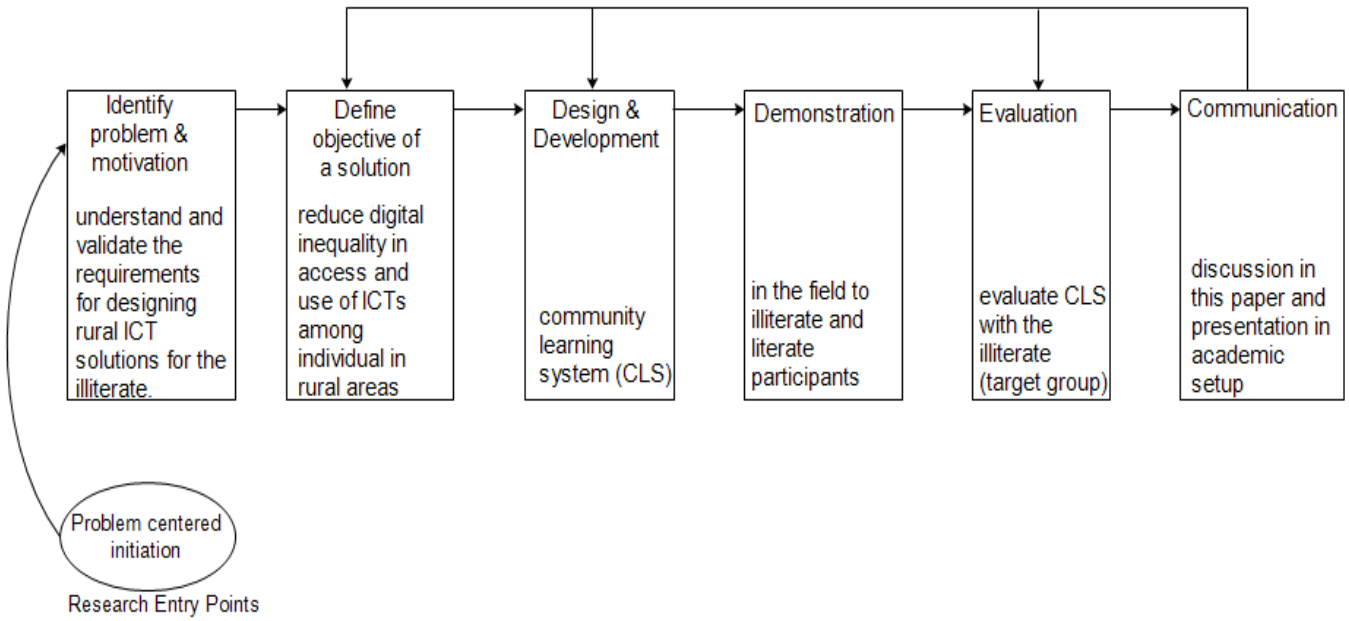


Fig. 4. DSR problem centered approach.

2) Problem identification and motivation: The existing ICT solutions in rural Kenya (for example websites or web applications or ICTs for information dissemination) are exclusively English and text-based in their user interfaces. Therefore, rendering it challenging for individuals with low literacy levels to access and utilize them. As a result, ICTs for relaying information to and in rural areas mainly target the educated and literate, excluding the illiterate and the semi-illiterate. Therefore, there is a need to understand and validate the requirements for designing rural ICT solutions which consider the illiterate people, allowing them to fully, on their own, participate in their access and use.

3) Objectives of the solution: The objective of this study is to validate the identified design needs for rural areas so as to inform solutions for enhancing ICT inclusivity in access and use among individuals. To achieve this, we developed an artifact which helped in identifying other special needs tailored to Kenyan rural areas. The results from artifact evaluation were used to develop a comprehensive and practical solution for designing rural ICTs in Kenya.

4) Design and development: Determining the intended functionality and architecture of the artifact is part of this activity, followed by the creation of the actual artifact [29]. Hence, prior to CLS creation, data was collected from the identified location of study, in accordance with [29] recommendations. Additionally, the existing practical approaches (in literature) for designing systems for the illiterate were considered. Study, [29] noted that the problems that have been identified may not directly correspond to the objectives for the artifact which was supported by our case, the artifact was disseminating information to the users but to us (the researchers), we were only using it to validate considerations for designing ICTs for the rural individuals especially the illiterate.

Now, to be explicit on CLS design and development process, the rural ICT design and development model was integrated with DSRM process, see Fig. 5.

a) Selection of location: The selection was guided by Fogg's argument that computers can be used to persuade people to change their behavior (doing something differently from the usual) as well as design ICT products which persuade people towards a behavior (adoption). The two villages (Cheleget and Kipgegei) were secretely practicing female genital mutilation and at the same time maize farming was no longer viable because of change in weather patterns. We assumed that applying Fogg frameworks in design of the artifacts was going to persuade them change their attitudes on certain vices and appreciate other farming innovations as advised by the agricultural extension officers like embracing drought resistant crops. With this, we were careful not to defiate from the objective of the study (validation). At thesame time, this formed part of our local content among those suggested by the participants.

b) Familiarizing with the community: On the first day, we went round the villages doing observations and familiarizing ourselves with the community's way of life where we interacted with thirthy one (31) individuals. We utilized the opportunity to share information about our study and the idea of creating CLS with them and or them. This was a way of creating awareness for the to-be designed ICT product. We explained to them the benefits which come with ICT interventions; including access to information which can help them improve their farming skills, business skills, health among others. We did not forget the fact that they could also get an opportunity to market their products through ICT solutions.

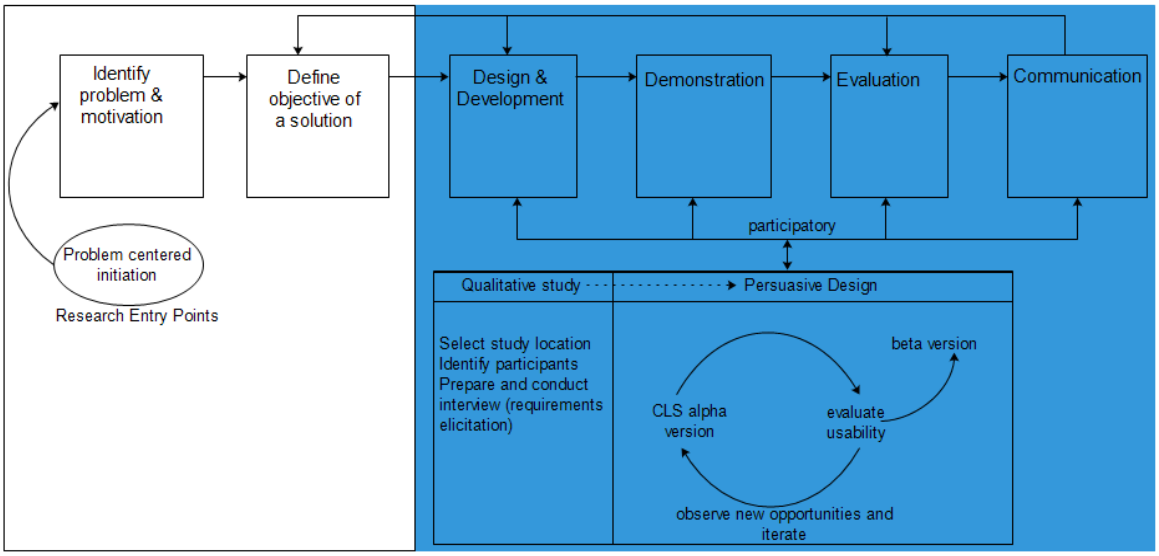


Fig. 5. DSRM-RICT integrated design and development model.

c) Selection of participants: Nine (9) participants were identified. They were willing to participate and support the study until the end. Four of them were promissing content creators. They were doing well in agriculture and interventions for food security in the village. For example, the farmer in Fig. 6 created content demonstrating how to build a beehive. This also took care of local content.

Fig. 6. Beehive construction demonstration at Cheleget village.

d) Preparing and conducting the interview (requirements elicitation): After we had familiarised ourselves and won the villagers for support, we conducted an interview. The survey asked the participants about their personal information such as age, familiarity with computers and the internet, educational background, and other relevant details. Then we sought to understand the things that would motivate them to access and use rural ICTs (computers and internet or community learning systems/portals), their abilities to use ICTs including literacy level, and the triggers that would prompt them to access and use the products. Another theme that we incorporated was the ICT product's / artifact's attributes as per resource and appropriation model [39]. We noticed that the question on attributes was too technical for the respondents, as they had not experienced these technologies before. In fact, one participant responded that he did not know how a computer looks like, and that if he had seen and used it then he could be having something to say about it. So, we decided at that point to embrace the saying that "seeing is believing." Therefore, we create a prototype and used it to probe their ICT design needs.

Overall, we interviewed twenty-seven (27) participants of whom, seventeen (17) were female and ten (10) were male. We stopped the interview at twenty-seventh individual because the interview was getting saturated. That is, response was being replicated and the only part that was changing was the demographics such as age, gender among other.

5) CLS development: After we had analyzed the requirements, including the local content, we immediately embarked on CLS development. All the requirements which had been recommended [12, 13, 15, 17, 19, 40, 41, 42, 43,] and indicated in integrated Fogg persuasive framework, Table VI, for example multisensory strategies [44] including audio, video were applied were implemented, following DSRM process. Fig. 7 is part of CLS's UI. Consistency was another consideration in CLS development. This is demonstrated in screenshots A, B, and C.

Fig. 7. CLS user interface.

Fig. 8 illustrates the CLS minimalistic approach [45] involving navigation (two-step process to accomplish a task). First step, a user clicks on a category of interest (button), for example, agriculture then a screen showing different items

(content) on this category appears. Second step, user selects content to read or watch/listen to the video. We presented information in local language, achieving user experience (UX) localization [46].

Fig. 8. Features of CLS.

Every button and icon was linked to an audio message. Anytime the mouse hovers on it, the message explained its purpose and the action to be accomplished if the user was to select it. For example, click on it if they were interested in that particular information. It enabled users to select the right information according to their preferences, minimizing errors at the same time enhancing self-efficacy.

6) CLS demonstration: The nine (9) participants were engaged in the usability evaluation of the prototype. This phase was crucial in understanding the user's design needs. Other three (3) educated and literate participants were approached and they agreed to give their feedback on the system. Since we had adhered to the developed framework in Table VI, the finding showed that the participants were satisfied, and were in agreement that the product was usable, particularly by the illiterate. Majority requested for additional content. Apart from the participants' responses, our observation was that those users who were unable to read or slow in reading, were unable to perceive the application's launching icon on their own. This was going against our goal of achieving users' self-efficacy (ability to complete a task without help). This feedback was then utilized for further refinements. Fig. 9 shows CLS modified launching icon which was more pronounced and visible.

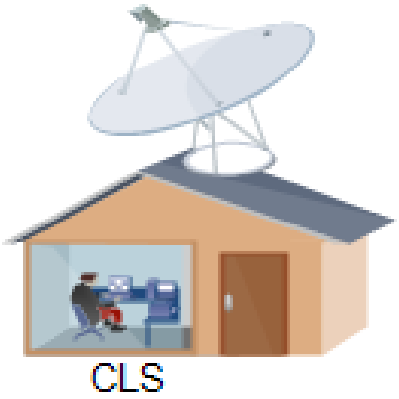


Fig. 9. CLS launching icon.

7) CLS evaluation: The beta version of CLS was evaluated for usability. On our initial visit, 31 people who expressed interest and were willing to take part in the study were specifically asked to participate in the evaluation. Among them is the group of nine people who participated in the design of CLS (community co-designers). Thirty-seven participants attended the session but only twenty-three (23) evaluated CLS. Though in attendance, the nine participants and five other literate people were excluded. At this stage the literate participants assisted in setting-up the room for evaluation and during participants' training session. We started the session by first training the participants on how to use CLS. Among the participants, 12 were male and the rest (11) were female. Fig. 10 shows the trainer projecting the steps (photo 1) and assisting the participants (photo 2). Majority of the participants were between the age of 25 and 60.

Fig. 10. CLS training session.

8) Communication: We discussed the findings of CLS evaluation in this paper and presented in an academic forum.

IV. RESULTS

We conducted a usability test to evaluate illiterate design needs, focusing on their ability to independently launch the system (self-efficacy). The timing period was determined during demonstration sessions where an average of the time taken by the literate and the illiterate participants was computed. During the evaluation, participants were asked to close (exit) CLS window, then locate and launch it within 180 seconds. Results in Table I revealed that 16 participants (70%) successfully located and launched the system on their first attempt, 3 participants (13%) succeeded after several attempts without assistance, and four participants (17%) required help after multiple unsuccessful trials.

TABLE I. TASK ONE- LAUNCHING CLS

Number of attempts	*Frequency (number of participants)*	*Percentage (%)*	*Average time (seconds)*
succeed on 1st attempt	16	70	48
succeeded after several attempts	3	13	123
failed after several trials	4	17	not successful

Task 2 evaluates users' self-efficacy by assessing their ability to independently complete a familiar task which involved opening a beehive making video. This task also provided

insights into their learnability and memory retention, as they had previously accessed the video multiple times. Allocating a duration of 120 seconds for this particular assignment, a total of 19 participants, which accounts for 83%, effectively managed to finalize it with an average completion time of 68 seconds. However, from Table II, four participants were unable to finish within the allotted time, mainly due to difficulties with mouse handling and positioning. Upon requesting a second attempt, half of these participants completed the task in an average of 63 seconds, while the remaining two still required assistance.

TABLE II. TASK TWO- FAMILIA TASK

Number of attempts	*Frequency (number of participants)*	*Percentage (%)*	*Average time (seconds)*
succeeded	19	83	68
failed after several attempts	4	17	elapsed time
2nd chance			
succeeded	2	50	63
failed after several attempts	2	50	not successful

In Task 3, participants were instructed to access a video on "Kalenjin Proverbs" which was uncommon task. However, our observations revealed that most participants typically viewed videos related to agriculture and health. As a result, the "Kalenjin Proverbs" video was not frequently accessed. During the training session, participants were informed that any information not related to agriculture or health was categorized under "general information." They were allocated 120 seconds to complete this task. Based on the data presented in Table III, it was observed that 16 individuals, constituting 70% of the total participants, effectively accomplished the task during their initial try, demonstrating an average duration of 69 seconds. Of the eight participants who failed initially, five succeeded on their second attempt, with an average time of 66 seconds. Despite multiple attempts, three participants were unable to complete the task.

TABLE III. TASK 3- UNCOMMON TASK

Number of attempts	*Frequency (number of participants)*	*Percentage (%)*	*Average time (seconds)*
succeeded	16	70	69
failed after several attempts	8	30	elapsed time
2nd chance			
succeeded	5	63	66
failed after several attempts	3	37	not successful

In Task 4, participants were instructed to close the CLS application, relaunch it, and then open content related to female genital mutilation (FGM). They were allocated 240 seconds to complete this task. As indicated in Table IV, despite the complexity of the task, 78% of participants successfully completed it. Upon a subsequent trial, 80% of the individuals successfully finished the assignment within an average duration of 128 seconds.

TABLE IV. TASK 4- LENGTHY TASK

Number of attempts	*Frequency (number of participants)*	*Percentage (%)*	*Average time (seconds)*
succeeded	18	78	186
failed after several attempts	5	22	elapsed time
2nd chance			
succeeded	4	80	128
failed after several attempts	1	20	not successful

We concluded the CLS evaluation session by asking the participants to verbally give their views on the system, as applied by [13]. This gauged their satisfaction and experience with CLS. All participants were using the computer and digital application for the first time. The system had employed the guidelines in Table VI, like use of video, audio cues, local language, and local language. Collected information included their engagement with CLS, its relevance (usefulness), their learning experience, their willingness to recommend it to the others and finally, challenges experienced while using CLS.

Finally, the study sought to understand rural users' willingness and ability to pay for the internet services. It started with a proposal to install CLS in their nearest market's cybercafé, located 8 kilometers from the village. The cybercafé was charging one shilling per minute for access and use of the internet. Results indicate that users were dissatisfied with the proposal, given the distance involved and monetary issue.

V. DISCUSSION

Similar to study [42] our study findings show that user self-efficacy was attained. This is explained by the task findings, 71%, 83%, 70% and 78% successfully, within the stipulated time and on their own, accomplished tasks 1, 2, 3 and 4 respectively.

Comparable to [12, 13], participants reported their overwhelming satisfaction with the CLS. They were satisfied with the use of local language and local content, noting that the language made it easy for them to learn and use the system while the local content matched with their information quest and needs. They found the content to be practical, and applicable to their day to day activities. They enjoyed using it and found it relevant (helpful/useful). Users appreciated the system's potential for learning, indicating that it could significantly enhance their knowledge and skills. Table V summarizes the impact of CLS on user experience, highlighting both positive aspects and challenges.

TABLE V. INTEGRATED FOGG PERSUASIVE FRAMEWORKS (FBM AND CAPTOLOGY)

Captology	*FBM Element*	*Design requirements/features*	*Guidelines*
Computer as a tool			
	Enhancing Ability	Increase self-efficacy (reduction of tasks) provide tailored information simplify/guide people through process (integrate training modules and audio aids) {minimalistic approach and UX localization}	Make navigation simple (few clicks) Design UI using less text (video and audio) present local content in local (mother) language Use intuitive images, symbols, and pictures Offering guidance through audio, video, and animated tutorials.
	Boosting Motivation	Provide immediate feedback	Information to reinforce positive behavior for example congratulates the user for listening to the video
	Effective Prompts	trigger decision making	use visual and audio cues like bright colors, icons, and sounds to guide users through tasks (uses visual/audio cues to prompt actions)
Computer as a medium			
	Enhancing Ability	Simulate cause and action	create interactive simulation that allow users to practice tasks or create a video of peers/expert simulating a task
	Boosting Motivation	Create engaging content Include demonstration	Incorporate stories of successful person for example in farming or business to motivate users or Show real-life scenarios where similar users successfully use the technology, enhancing the perceived relevance and usefulness.
	Effective Prompts	Provide interactive guides provide prompts during the simulation.	Use animated characters or avatars to provide step-by-step instructions and encouragement and Trigger actions within the simulation to guide users towards desired behaviors.
computer as a Social Actor			
	Enhancing Ability	Provide social support Incorporate mentorship roles	During training, allow users to practice on their own (peer learning) The application should allow interested community members for example farmers to share their experiences. Incorporate virtual mentors or coaches that provide guidance and support.
	Boosting Motivation	Provide social proof Include recognition	Design with the community and let them be content creators Shows successful peer examples and their achievements Show that others in the community are using the technology, leveraging the influence of social norms. Publicly recognize achievements within the application, such as leaderboards or community shout-outs.
	Effective Prompts	Apply social prompt Consider community engagement	Sends reminders from community social workers, agriculture extension officer, community leaders among others. Use social challenges or group tasks to motivate users through collective action.

Due to the cost and location, rural users expressed their unwillingness to participate in its access and use. The participants expressed their dissatisfaction with the proposed location terming it as far away, requiring additional time, cost and effort to travel. Largely, the economic constrain among individuals in rural areas significantly impacted their willingness to pay for internet services. Data and observations showed that the majority of users had limited financial resources. Majority struggled to afford basic necessities, including school fees for their children and daily living expenses. Given their low income levels, the users were not willing to allocate funds for accessing the internet services. The idea of paying for information was deemed impractical and unaffordable. Similar to study [13], all the participants positively articulated their willingness to recommend CLS to the other members of the community.

TABLE VI. RATING CLS SATISFACTION

Positive aspects	***views***
Engagement	Participants enjoyed the interactive elements provided by the video and audio cues. These features made the system engaging and easy to navigate.
Relevance	The use of local language and content resonated well with the users thus made the information more accessible and meaningful.
Learning Potential	Users felt that the system was a valuable educational tool with a very high potential of facilitating significant learning and personal development.
challenges	
Mouse Handling	participant's difficulty in handling the mouse was the primary and most notable challenge to be reported. Many found it challenging to use. This indicated a steep learning curve which was experienced as a result of this input device.
Learning Time	Participants felt they needed more time to become proficient with the system especially in relation to mouse handling. They suggested that if alternative input methods were provided then their experience could improve.
Content	Participants felt CLS needed to be populated with more content particularly on how to grow and manage various crops.

VI. CONCLUSION

This study validated design requirements for rural areas, focusing on the illiterate. It followed DSRM process and applied Fogg integrated frameworks. The artifact that was used for validation, applied the specified design features and guidelines indicated in the framework. In 2022, [32] proposed the same requirements for creating systems designed to address gender inequality. Furthermore, this approach is capable of generating intuitive artifacts, closing the digital gap not only between illiterate and literate individuals, but addresses the gender and age divide as well. We found out that user experience localization enhances user's engagement with the ICT artifacts or products. This was achieved with the use of local language and local content. It was ascertained [46] that adaptation of UX enables an application to bridge divides among various demographics. Content was development in local language, meaning that there was no need for interpreters or difficulty in comprehension. At the same time, tailored information was provided through CLS as the community members were fully involved in the design and development of CLS. This means that the information was relevant to users' needs and their context. Furthermore, the local content was based on users' demand. Localization strategy showed full potential of encouraging ICT inclusivity, fostering user engagement with the ICTs.

Application of videos and audio cues significantly impacted participant's UX with CLS. It made CLS intuitive and easy to learn. According to study [44] combining various techniques for information dissemination or learning is known as multisensory approach. Our study revealed that multisensory approach is applicable in the design of rural ICTs (artifacts). Furthermore, CLS applied minimalistic design principles [46] which included few clicks and intuitive navigation, enhancing user's self-efficacy. This finding agrees with [6], where they revealed that the act changes users' attitude and behavior by exposing them to information which may have not been readily available at their disposal, hence they may be persuaded to apply the learned information in their daily lives.

This study maximized on content categorization. This strategy involved putting together related information helping the users to locate information with ease, and minimizing the time taken to search for information. The results highlight the varying levels of self-efficacy among participants and the need to carefully craft solutions for special groups like those with low cognitive levels, low technical skills and low confidence. In their study [47] found this to be the case as they concluded that ICT user interface and content designed for individuals with low literacy skills should consider factors beyond just illiteracy, including additional cognitive differences among these users.

This study reveals the positive significance of peer participation in the design and acceptability of ICTs by rural people. Peers were involved in the design, evaluation and content creation. Also, during training, peer learning was encouraged. They patronized the content created by their own village mates than others. This enhanced learning speed and made the participants engaged with the system. Additionally, other input devices need to be explored for the users with low technical skills like dragging and positioning the mouse.

DSRM provided significant benefits for our research endeavor. Its structured approach ensured thorough investigation and validation of the specific needs of our target demographic, while its iterative nature allowed for continual refinement based on real world feedback. This adaptability was crucial for addressing the diverse challenges faced by illiterate users in rural areas. Furthermore, the focus of the methodology on practical implementation harmonized effectively with our objective of establishing a community learning system that was accessible and user-friendly. Additionally, DSR's efficient use of resources made it a time effective and economical choice, allowing us to maximize our impact within the constraints of our research budget. Thus DSR provided the ideal framework for developing inclusive ICT solutions that meet the specific needs of rural populations.

Additionally, this study revealed that combination of DSR and Fogg's persuasive frameworks helps in creating usable and accessible ICT solutions customized to the needs of the target population. It enhanced the effectiveness of ICT solutions for illiterate individuals. These approaches had a significant impact in the achievement of user self-efficacy.

Despite the diffusion of mobile phones, majority of the individuals in rural areas have not switched to internet enabled phones (smartphones), [31, 48]. According to study [49,50] featured phones are still common in rural areas because of prevalence in illiteracy, affordability concerns, long battery life and weak network connections Therefore similar to [19], CLS was a computer based system.

Hence, our investigation and validation showed that critical ICT design needs for rural users include the use of local language, local content, videos, audio, few clicks, touch screen, noise-proof devices, content categorization, peer collaboration, public RICTs and low cost or subsidized ICT resources.

Additionally, location is also important in access and use of RICTs. The design process should be participatory.

We conclude this study by noting that design solutions targeting vulnerable demographics is key to the success of designing for equality. In other words, that which works or usable by the aged, illiterate, uneducated and the poor works best or more usable for the young, literate, educated, and the rich (financially stable). Thus enhancing access and use of rural ICTs.

VII. Recommendation and Future Work

This study recommends the application of DSRM-RICT integrated design and development model and the Fogg frameworks. These methodologies are critical in the design and development of ICT artifacts with enhanced user experience.

CLS was computer based system. Therefore, following the DSRM-RICT integrated design and development model, together with integrated Fogg persuasive frameworks, future researchers are recommended to design and evaluate rural artifacts for mobile phones.

Also, future studies on rural informatics can look at various multisensory methodologies which can foster user engagement with the ICTs, for example apart from auditory (audio) and visual (video), they can consider haptic (tactile feedback).

Additionally, rural informatics researchers are encouraged to create interactive simulations which will enable users to practice a task as of real world.

Finally, this study observation is, however usable the artifact is, users may decline to access it if the location is not favorable. FBM [4] asserts that ability is comprised by anything that makes it difficult to perform a task. This in turn influences motivation in a negative manner, rendering the triggers (awareness) meaningless. Therefore, for accessibility purposes, this paper recommends strategic placement of rural ICTs.

References

[1] Vassilakopoulou, P., & Hustad, E. (2021). Bridging Digital Divides: A Literature Review and Research. *Information Systems Frontiers*, 1-15.

[2] UN-HABITAT. (2021). *Assessing the Digital Divide: Understanding internet connectivity and digital literacy in cities and communities.* Nairobi: United Nations Human Settlements Programme (UN-Habitat).

[3] Van Deursen, A., & van Dijk, J. (2019). The first-level digital divide shifts from inequalities in physical access to inequalities in material access. *New Media & Society, 21*(2), 354-375.

[4] Fogg, B. J. (2003). *Persuasive Technologies: Using computers to change what we think and do.* United states of America: Morgan Kaufmann.

[5] Wenker, K. (2022). A systematic literature review on persuasive technology at the workplace. *Patterns, 3*(8), 1-9.

[6] Fogg, B. J., Cuellar, G., & Danielson, D. (2007). Motivating, influencing, and persuading users. In J. A. Jacko, A. Sears, J. A. Jacko, & A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications.* (2nd edition ed., pp. 133-144). New York: Lawrence Erlbaum Associates.

[7] Mahesh, S. (2023, August 16). *User-Centric Alchemy: Transforming Design with BJ Fogg's Behavior Model*. Retrieved from Medium: https://medium.com/@saisriramm/user-centric-alchemy-transforming-design-with-bj-foggs-behavior-model-36bc895cfc42

[8] Julius, G. (2023, October 24). *Using The Fogg Behaviour Model to Create a More User-Centric Digital olutions*. Retrieved from iPF Softwares Team Observing VSLA Users using Chomoka Mobile Application : https://www.linkedin.com/pulse/using-fogg-behaviour-model-create-more-user-centric-digital-julius-mqu6f

[9] Mallawaarachchi, S. R., Tieppo, A., Hooley, M., & Horwood, S. (2023). Persuasive design-related motivators, ability factors and prompts in early childhood apps: A content analysis. *Computers in Human Behavior, 139*.

[10] Sarwar, A., Mukhtar, H., Maqbool, M., & Belaid, D. (2015). SmartFit: A Step Count Based Mobile Application for Engagement in Physical Activities. *International Journal of Advanced Computer Science and Applications*, 271-278.

[11] Filippou, J., Cheong, C., & Cheong, F. (2019). Evaluating Persuasion in a Digital Learning Environment. *ACIS 2019 Proceedings* (pp. 516-526). Perth Western Australia : AIS Electronic Library.

[12] Alfayez, Z. H. (2022). Design and implement a mobile-based system to teach traffic signs to illiterate people. *i-com, 21*(3), 353–364.

[13] Islam Nazrul, M., Ahmed, A. M., & Islam, N. A. (2020). Chakuri-Bazaar: A Mobile Application for Illiterate and Semi-Literate People for Searching Employment. *International Journal of Mobile Human Computer Interaction, 12*(2), 22-39.

[14] Mtika, P., & Abbott, P. (2023). Literacy and quality of life: a study of adults with poor literacy capabilities in western Rwanda. *A journal of comparative and international education* , 1-17.

[15] Sadeeq, J., Imran, M., Iftikhar, A., Majid, A., & Fazal, Q. K. (2019). A Systematic Feasibility Analysis of User Interfaces or Illiterate Users. *Physical and Computational Sciences, 56*(4), 75–91.

[16] UNESCO. (2020). *UNESCO Ralies international organizations, civil society and private sector partners in a broadband coalition.* UNESCO.

[17] Islam, M. N., Khan, N. I., Inan, T. T., & Sarker, I. H. (2023). Designing User Interfaces for Illiterate and Semi-Literate Users: A Systematic Review and Future Research Agenda. *Sage Open, 13*(2).

[18] Oinas-Kukkonen, H., & Harjumaa, M. (2009). Persuasive Systems Design: Key Issues, Process Model, and System Features. *Communications of the Association for Information Systems*, pp-pp.

[19] Rahman, A., & Fukuda, A. (2015). User Interface Design of E-Learning System for Functionally Illiterate People. *International Journal of Advanced Computer Science and Applications, 6*(11), 126-134.

[20] Teperi, A.-M., Gotcheva, N., & Aaltonen, K. (2021). Design thinking perspective for developing safety management practices in nuclear industry. *Human Factors in the Nuclear Industry*, 309-326.

[21] Benda, N. C., Montague, E., & Valdez, R. S. (2020). Design for inclusivity. In S. Arathi, & S. Farzan, *Design for Health* (pp. 305-322). United States: Academic Press.

[22] Kapuire, G., Winschiers-Theophilus, H., Blake, E., Bidwell, N. & Chivuno-Kuria, S. (2010). A revolution in ICT, the last hope for African Rural Communities Technology Appropriation. *IDIA*, Cape Town.

[23] Siew, S., Yeo, A. W., & Zaman, T. (2013). Participatory Action Research in Software Development: Indigenous Knowledge Management Systems Case Study. M. Kurosu (Ed.): Human-Computer Interaction, *Springer-Verlag Berlin Heidelberg Part I, HCII 2013, LNCS 8004*, 470–479.

[24] Schaffers, H., Guzman J. G. & Merz, C. (2008). *An Action Research Approach to Rural Living Labs Innovation.* In: P. Cunningham and M. Cunningham (Eds), *Collaboration and knowledge Economy: Issues, Applications, Case Studies*. IOS press, 617-624.

[25] Doerflinger, J. & Gross, T. (2010). Technical ICTD - A User Centered Lifecycle. A. Pont, G. Pujolle, and S.V. Raghavan (Eds.): *WCITD/NF 2010, IFIP AICT*, 327, 72–83.

[26] Davis, J. (2009). Design Methods for Ethical Persuasive Computing. *Proceedings of the 4th International Conference on Persuasive Technology* (pp. 1-8). Claremont, California, USA: ACM.

[27] Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a Framework for Evaluation in Design Science Research. *European Journal of Information Systems, 25*, 77-89.

[28] Premananda, G.-A. I., Tjahyanto, A., & Mukhlason, A. (2022). Design Science Research Methodology and Its Application to Developing a New Timetabling Algorithm. *Proceedings - 2022 IEEE International Conference on Cybernetics and Computational Intelligence, CyberneticsCom 2022* (pp. 433-438). Malang, Indonesia: Institute of Electrical and Electronics Engineers Inc.

[29] Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems, 24*(3), 45-77.

[30] Bilandzic, M., & Venable, J. (2011). Towards participatory action design research: adapting action research and design science research methods for urban informatics. *Journal of Community Informatics, 7*(3), 1-23.

[31] Okello, F. (2024). Bridging Kenya's Digital Divide: Context, Barriers and Strategies. *Digital Policy Hub. Centre for International Governance Innovation*, 1-13. Retrieved from https://www.cigionline.org/static/documents/DPH-Paper-Okello.pdf

[32] UN-Women. (2022). *Gender analysis in technical areas: Digital Inclusion.* New York: UN System Coordination Division of UN Women

[33] Kuteesa, K. N., Akpuokwe, C. U., & Udeh, C. A. (2024). Gender Equity in Education: Addressing Challenges and Promoting Opportunities for Social Empowerment. *International Journal of Applied Research in Social Sciences, 6*(4), 631-641.

[34] Mustafa, F., Nguyen, H. T., & Gao, X. (2024). The challenges and solutions of technology integration in rural schools: A systematic literature review. *International Journal of Educational Research, 126*. doi:https://doi.org/10.1016/j.ijer.2024.102380

[35] St-Maurice, J., Burns, C., & Wolting, J. (2018). Applying Persuasive Design Techniques to Influence Data-Entry Behaviors in Primary Care: Repeated Measures Evaluation Using Statistical Process Control. *JMIR Hum Factors, 5*(4).

[36] Shafin, N. A., & Abdullah, N. H. (2018). Persuasive System Practices in Mobile Application Development. *ournal of Physics: Conference Series, 1049*, 1-15.

[37] Won, J. (2018). PSPB (Persuasive Service Design Strategies Based on the Theory of Planned Behavior) Methodology for User Behavior Modification. *Archives of Design Research, 31*(3), 17-31.

[38] Alduhailan, G., & Alshamari, M. (2016). Influence of Adopting a Text-Free User Interface on the Usability of a Web-based Government System with Illiterate and Semi-Literate People. *International Journal of Advanced Computer Science and Applications, 7*(8), 181-188.

[39] Van-Dijk, J. (2020). *Closing the Digital Divide. The Role of Digital Technologies on Social Development, Well-Being of All and the Approach of the Covid-19 Pandemic.* Retrieved from Conference: Virtual Expert Group UN Meeting on "Socially just transition towards sustainable development: The role of digital technologies on social development and well-being of all": https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/07/Closing-the-Digital-Divide-by-Jan-A.G.M-van-Dijk-.pdf

[40] Belay, E. G., McCrickard, S. D., & Besufekad, S. A. (2016). Mobile User Interaction Development for Low-Literacy Trends and Recurrent Design Problems: A Perspective from Designers in Developing Country. In P. Rau, *Cross-Cultural Design* (pp. 409-417). Switzerland: Springer International Publishing.

[41] Barclay, P. A., & Bowers, C. A. (2017). Design for the Illiterate: A Scoping Review of Tools for Improving the Health Literacy of Electronic Health Resources. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61*(1), 545-549.

[42] Ahmed, A. M., Islam, M. N., Jannat, F., & Sultana, Z. (2019). Towards Developing a Mobile Application for Illiterate People to Reduce Digital Divide. *International Conference on Computer Communication and Informatics* (pp. 23 – 25). Coimbatore, INDIA: ICCCI -2019.

[43] Volpe, G., & Gori, M. (2019). Multisensonsory interactive technologies for primary education: From science to technology. *Frontiers in Psychology, 10*, 1-8.

[44] Vernon, B. (2023, August 07). *Exploring the Principles of Minimalism in UI/UX Design*. Retrieved from Medium: https://brianvernon1.medium.com/exploring-the-principles-of-minimalism-in-ui-ux-design-532093a083ad

[45] Optimational. (2023, August 14). *UX Localisation: Best Practices for a Multilingual Website*. Retrieved from Optimational: https://www.optimational.com/blog/ux-localisation-best-practices/

[46] Medhi, I., Menon, R. S., Cutrell, E., & Toyama, K. (2010). *Beyond Strict Illiteracy:Abstracted Learning Among Low-Literate Users.* INternational Conference of Information and Communication Technologies and Development. Retrieved from http://cutrell.org/papers/ICTD2010%20Medhi%20et%20al.pdf

[47] Malhotra, A. (2024 , May 06). *Bridging Kenya's digital divide calls for inclusive development*. Retrieved from Business Daily: https://www.businessdailyafrica.com/bd/opinion-analysis/columnists/bridging-kenya-s-digital-divide-calls-for-inclusive-development--4615080

[48] Kabui, M. (2024, 07 22). *More Kenyans struggle to buy phones for Internet access*. Retrieved from Business Daily: https://www.businessdailyafrica.com/bd/economy/more-kenyans-struggle-to-buy-phones-for-internet-access-4697856#google_vignette

[49] Umidha, S. (2023, 03 09). *Feature phones still dominate Kenya's handset market in a smartphone age*. Retrieved from Financial Fortune: https://www.financialfortunemedia.com/feature-phones-still-dominate-kenyas-handset-market-in-the-smartphone-age/

[50] Kapondera, S. K., & Mtambo, J. (2023). Why do people use telecentres in the age of mobile technologies? Answers from Malawi. *The electronic jornal of information systems in developing countries, 90*(2).

Brain Tumor Segmentation of Magnetic Resonance Imaging (MRI) Images Using Deep Neural Network Driven Unmodified and Modified U-Net Architecture

Nunik Destria Arianti[1], Azah Kamilah Muda[2]*
Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia[1, 2]
Department of Information System, Nusa Putra University, Sukabumi, Indonesia[2]

***Abstract*—Accurately separating healthy tissue from tumorous regions is crucial for effective diagnosis and treatment planning based on magnetic resonance imaging (MRI) data. Current manual detection methods rely heavily on human expertise, so MRI-based segmentation is essential to improving diagnostic accuracy and treatment outcomes. The purpose of this paper is to compare the performance of detecting brain tumors from MRI images through segmentation using an unmodified and modified U-Net architecture from deep neural network (DNN) that has been modified by adding batch normalization and dropout on the encoder layer with and without the freeze layer. The study utilizes a public 2D brain tumor dataset containing 3064 T1-weighted contrast-enhanced images of meningioma, glioma, and pituitary tumors. Model performance was evaluated using intersection over union (IoU) and standard metrics such as precision, recall, f1-score, and accuracy across training, validation, and testing stages. Statistical analysis, including ANOVA and Duncan's multiple range test, was conducted to determine the significance of performance differences across the architectures. Results indicate that while the modified architectures show improved stability and convergence, the freeze layer model demonstrated superior IoU and efficiency, making it a promising approach for more accurate and efficient brain tumor segmentation. The comparison of the three methods revealed that the modified U-Net architecture with a freeze layer significantly reduced training time by 81.72% compared to the unmodified U-Net while maintaining similar performance across validation and testing stages. All three methods showed comparable accuracy and consistency, with no significant differences in performance during validation and testing.**

Keywords*—Accuracy; brain tumor; DNN; U-Net architecture; comparison performance

I. INTRODUCTION

Brain tumors are among the most life-threatening conditions, with the potential to drastically affect the quality of life and survival of patients. Early and accurate diagnosis is paramount to developing effective treatment strategies, often involving surgery, radiotherapy, or chemotherapy. Magnetic resonance imaging (MRI) is widely recognized as one of the most accurate non-invasive imaging techniques for detecting brain abnormalities. It provides detailed views of brain structures and tissue compositions, allowing clinicians to assess tumor size, location, and growth patterns from research in quantitative MRI [1], research review [2], and discussion from Advancements in Neuroimaging to detect brain tumors [3]. However, interpreting these MRI images is complex and time-consuming, requiring a high level of expertise [4]. As a result, manual segmentation of brain tumors can be inconsistent, subjective, and prone to human error. Automated segmentation methods using advanced machine learning and deep learning techniques have been explored to address these challenges in recent years using image cardiac radiography [5], image Alzheimer's disease in the human brain [6], and brain tumor detection using MRI [7].

Deep neural networks (DNNs), particularly convolutional neural networks (CNNs), have shown exceptional promise in image analysis tasks, including medical image segmentation. CNNs are highly effective at identifying patterns in image data by leveraging layers that automatically learn relevant features such as edges, textures, and complex structures. Among CNN-based methods, the U-Net architecture has become a leading tool for medical image segmentation due to its robust performance. Initially designed for biomedical image segmentation, U-Net uses an encoder-decoder structure, where the encoder extracts features from the input image, and the decoder uses those features to reconstruct segmented regions [8, 9]. U-Net's symmetrical architecture enables precise localization of structures, making it especially suitable for segmenting brain tumors. According to Yousef, et al. [10], despite its strengths, the standard U-Net architecture can be prone to overfitting, mainly when training on small datasets like MRI brain scans, leading to decreased generalization to unseen data.

Several improvements have been proposed to address the limitations of the original U-Net architecture, focusing on enhancing the model's generalization and stability during training. Two commonly employed techniques, batch normalization and dropout, have been used to learn parameter effects for segmenting brain tumors [11], mandible bones [12], and brain tumors using U-Net [13]. Batch normalization is used to normalize the input to each layer during training, reducing internal covariate shifts and accelerating the learning process. By stabilizing the learning process, batch normalization helps improve the performance of deep neural networks, mainly when applied to complex datasets like MRI images. On the other hand, dropout is a regularization technique that randomly disables a fraction of neurons during training, forcing the network to learn more robust features. This helps to mitigate overfitting by preventing the model

from relying too heavily on specific neurons [14]. Batch normalization and dropout can significantly enhance the U-Net architecture's ability to generalize from training data to new, unseen cases.

In addition to these regularization techniques, another modification involves freezing specific layers of the U-Net architecture during training. The freeze layer method limits the number of trainable layers, essentially "freezing" network parts to prevent them from being updated during training. This can be particularly useful when fine-tuning a model on a smaller dataset or using transfer learning [15]. By restricting the training depth, the model is less likely to overfit the specific training data and can generalize better to new data. Furthermore, freezing layers can reduce the computational cost of training, as fewer parameters need to be updated [16]. This makes it an attractive option for optimizing the training process without sacrificing performance. The combination of batch normalization, dropout, and freeze layers has the potential to significantly improve the efficiency and accuracy of brain tumor segmentation using MRI images.

The U-Net architecture, in original and with modifications in architecture, remains one of the most widely used models for medical image segmentation tasks. The original U-Net is effective but may struggle with specific challenges inherent to brain tumor segmentation, such as distinguishing between subtle boundaries of healthy and tumorous tissues [8]. Incorporating batch normalization and dropout into the encoder layers enhances the model's ability to avoid overfitting, allowing it to learn more generalized features. Furthermore, introducing a freeze layer can optimize computational efficiency, making the model more robust and faster to train. These architectural improvements hold great promise for improving the segmentation of brain tumors, ultimately contributing to better diagnostic accuracy and treatment planning.

This paper compares the performance of brain tumor segmentation using unmodified and modified U-Net architectures driven by deep neural networks (DNN). Specifically, two modifications are examined: (1) batch normalization and dropout added to the encoder layer, and (2) a freeze layer introduced to limit the depth of the training process. We selected the modified U-Net architecture with batch normalization, dropout, and freeze layers due to its effectiveness in addressing specific challenges in brain tumor segmentation. U-Net, widely used in medical image segmentation, excels at delineating complex structures, but it can struggle with overfitting, particularly on small datasets like MRI. To mitigate this, batch normalization stabilizes training by normalizing layer inputs, while dropout prevents over-reliance on specific neurons, reducing overfitting. Additionally, freeze layers limit trainable parameters, improving generalization and reducing computational cost. Standard methods often fail to handle subtle tumor boundaries and are prone to overfitting with small datasets. Our modifications enhance the model's robustness, efficiency, and accuracy in segmenting brain tumors. These improvements address limitations in existing approaches, making our method more suitable for the complexities of brain tumor segmentation from MRI images, where precise and reliable detection is critical.

The remainder of this paper is structured as follows: Section II overviews brain tumor segmentation using deep learning as a related work. Section III explains the dataset and methods in detail. In Section IV, we compare and analyze experiments and discussions from our studies. Finally, Section V concludes and describes future work.

II. Related Work

Research on brain tumor segmentation using deep learning has gained substantial attention in recent years due to the increasing availability of medical imaging data and the advancements in computational methods. Early approaches to brain tumor segmentation relied heavily on traditional image processing techniques such as thresholding, region growing, and edge detection [17-19]. While these methods could detect certain features, they lacked the sophistication needed to capture brain tumors' complex and heterogeneous nature. As a result, their accuracy was often limited, especially when dealing with tumors with irregular shapes or indistinct boundaries. The development of machine learning models, particularly convolutional neural networks (CNNs), revolutionized this field by enabling more automated, accurate, and efficient segmentation techniques, especially in medical imaging applications such as brain tumor detection.

One of the first major breakthroughs in deep learning-based medical image segmentation came with the introduction of the U-Net architecture by Ronneberger, et al. [20]. Initially designed for biomedical image segmentation tasks, U-Net rapidly gained popularity due to its simplicity and effectiveness. The architecture's encoder-decoder structure enables precise segmentation by combining high-resolution features from the encoder with upsampled features in the decoder. U-Net has been successfully applied to various medical imaging tasks, including brain tumor segmentation, and has become a standard baseline model in the field. However, despite its success, U-Net often struggles with overfitting and performance stability, mainly when dealing with minor or imbalanced datasets, which are common in medical imaging.

Various enhancements to the U-Net architecture have been proposed to address these issues. Isensee, et al. [21] introduced modifications such as deep supervision and residual connections to improve the network's ability to learn from complex medical imaging data. Other works have focused on regularization techniques like batch normalization and dropout to enhance model stability and generalization. Batch normalization, introduced by Ioffe [22], is commonly used to normalize the inputs to each layer, speeding up training and reducing internal covariate shifts. Dropout, proposed by Srivastava, et al. [23], helps prevent overfitting by randomly deactivating neurons during training, forcing the model to learn more generalized features. Various CNN architectures, including U-Net, have widely adopted these techniques to improve their performance on medical image segmentation tasks.

Additionally, several researchers have explored transfer learning and fine-tuning in medical image segmentation. Transfer learning allows a pre-trained model to be adapted to a new task by freezing specific layers during training and updating only a subset of parameters. In brain tumor segmentation, transfer learning can be particularly beneficial when working with small datasets, as it enables models to leverage knowledge from larger, more general datasets. Kamnitsas, et al. [24] demonstrated the effectiveness of transfer learning for brain lesion segmentation using a 3D CNN, achieving better performance than training from scratch. The use of freeze layers, which limit the depth of training, has also been explored to reduce overfitting and optimize computational efficiency, particularly in resource-constrained environments.

Moreover, ensemble learning methods have been explored to improve segmentation performance further. DeepMedic, introduced by Kamnitsas, et al. [25], combines multiple CNNs to form an ensemble model, resulting in more robust and accurate brain tumor segmentation. Other works have integrated different architectures, such as residual networks and attention mechanisms, to enhance segmentation models' feature extraction and localization capabilities. In summary, the field of brain tumor segmentation has seen significant advancements, with numerous modifications to the original U-Net architecture and the incorporation of advanced techniques such as batch normalization, dropout, freeze layers, and ensemble learning approaches. These improvements enhance model stability, reduce overfitting, and increase segmentation accuracy.

III. Dataset and Methods

A. Data Sets

The dataset used in this study is a public 2D brain tumor dataset authored by Cheng, et al. [26], Cheng, et al. [27]. The dataset contains 3064 T1-weighted contrast-enhanced images with three kinds of brain tumors (meningioma with 708 slices, glioma with 1426 slices, and pituitary tumor with 930 slices) from 233 patients that were scanned by magnetic resonance imaging (MRI), all with ground truth segmentations of the tumors. It was randomly split into a training set of 2451 images, an internal validation set of 306 images, and a testing set of 307 images.

B. The Proposed Method

In this study, three architectures from deep neural network (DNN) were used, including (1) an unmodified U-Net architecture, (2) a modified U-Net architecture with batch normalization and dropout on the encoder layer without the freeze layer, and (3) a modified U-Net architecture with batch normalization and dropout on the encoder layer with a freeze layer. The U-Net architecture added with freeze layers aims to limit the training depth by skipping certain layers when training the model.

C. Evaluation of the Models Performance

During the calibration stages (training and internal validation) for the three architectures examined in this study, the performance of the model development is evaluated through training loss and validation loss. Moreover, as a widely-used evaluation metric in object detection and image segmentation tasks, the intersection over union (IoU) overlap between predicted bounding boxes and ground truth boxes is also evaluated with scores ranging from 0 to 1. After that, the model is tested using a dataset of 307 images.

To find out more details regarding the model performance of each stages (training, internal validation, testing), 10 images were randomly taken from each stages to be compared to the ground truth segmentations of the tumors. Each image will be checked for the number of true positive, true negative, false negative, and false positive variables to calculate precision, recall, f1-score, and accuracy. After that, the average with the standard deviation for the three architectures was compared through the ANOVA statistical test to determine the model performance's significance ($p<0.05$). Duncan's multiple range test will be carried out if at least one difference exists between the three architectures.

IV. Result and Discussions

A. Unmodified U-Net Architecture

Fig. 1(a) shows the training history for 35 epochs using unmodified U-Net architecture for the loss training (0.1453±1.21E-01) and internal validation (0.1931±5.10E-02) stages. There is a loss gap of around 32.94% between training and validation. The loss curve for validation shows that the best loss can be achieved at an epoch of 35. The time calculation for running training for this method is 197.11±0.37s. The performance of the model by the intersection of union (IoU) on Fig. 1(b) for unmodified U-Net architecture is 0.7376±1.41E-01 in training and 0.6861±2.92E-02 in validation. There is a difference in IoU of approximately 6.98% between the training and validation results.

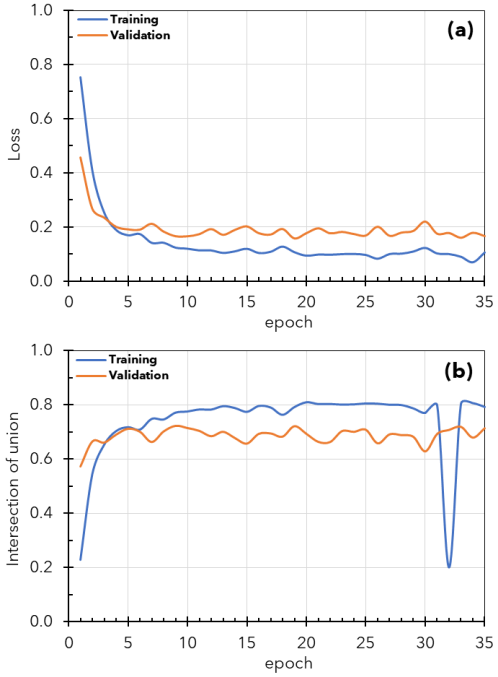


Fig. 1. Training and validation history of 35 epochs iteration from unmodified U-Net architecture model (a) accuracy and (b) IoU.

Fig. 2 presents the model's performance when sampling 10 images for the confusion matrix. The model from unmodified U-Net is more accurate (Fig. 2(a)) than the validation and testing stages except in images number 4, 5, and 7 (see Fig. 3). The average accuracy at the training, validation, and testing steps were 0.9977+7.46E-04, 0.9979+7.27E-04, and 0.9979+1.18E-03, respectively. The ANOVA test results for the accuracy parameters at the training, validation, and testing steps were not significantly different (p>0.05), indicating that the model is steady and there is no under/overfitting.

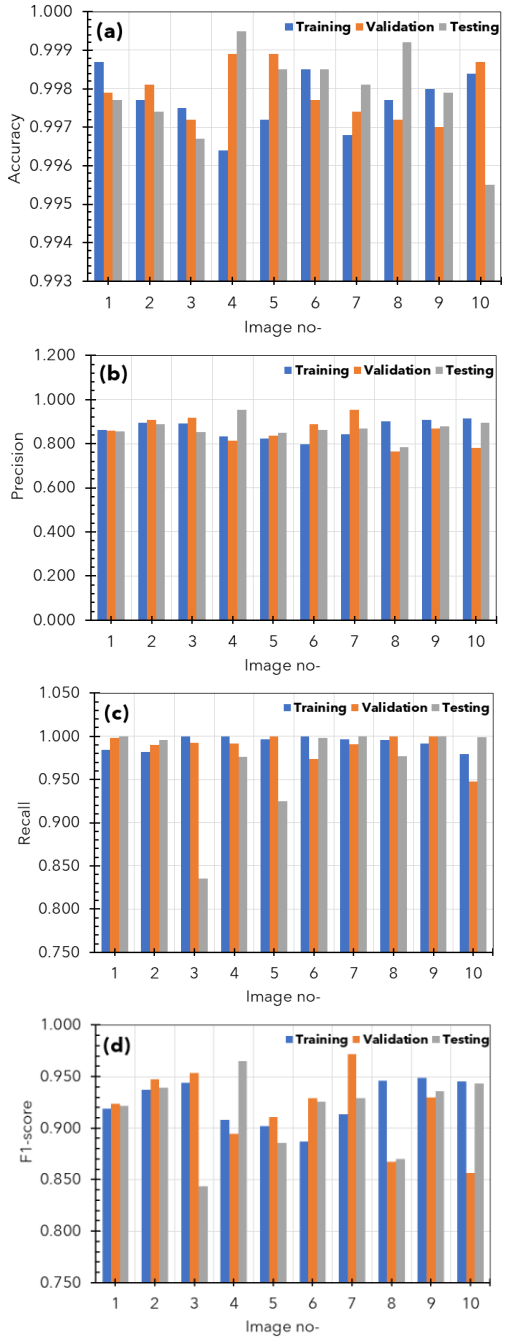


Fig. 2. Performance of sampling image from unmodified U-Net in a confusion matrix (a) accuracy, (b) precision, (c) recall and (d) f1-score.

Precision can be used as a parameter to measure how consistent the model is in predicting the quality of positive predictions. The average precision (see Fig. 2(b)) at the training, validation, and testing stages were 0.8670+4.12E-02, 0.8593+6.04E-02, and 0.8685+4.27E-02, respectively. The statistical test results using ANOVA did not find any significant differences between the three stages (p>0.05). This indicates that the results obtained through precision did not find any false patterns from the tested dataset.

Next, the recall parameter can show the total number of actual positive cases that are correctly predicted to show the sensitivity of the model. The average recall (Fig. 2(c)) at the training, validation, and testing stages are 0.9927+7.95E-03, 0.9884+1.63E-02, and 0.9706+5.30E-02, respectively. The ANOVA statistical test results revealed no significant differences across the three stages (p>0.05), indicating that the recall results did not detect incorrect patterns in the tested dataset.

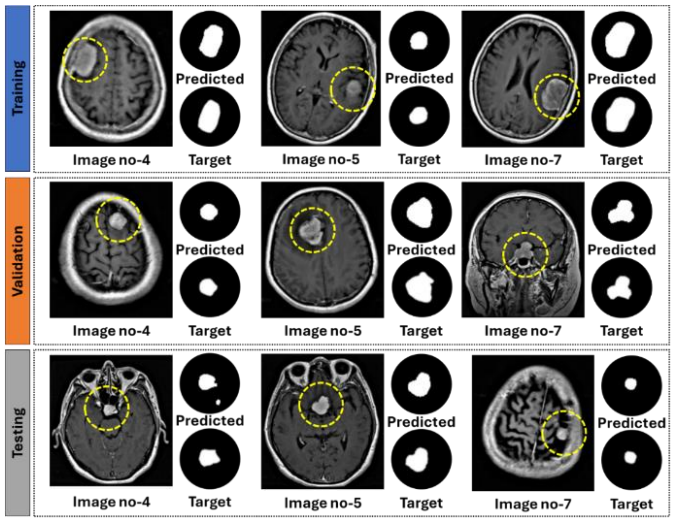


Fig. 3. Details of sample images 4, 5, and 7 in training, validation, and testing by unmodified U-Net architecture model.

B. Modified U-Net Architecture without Freeze Layer

The training and validation history and IoU for 35 epochs using modified U-Net architecture without a freeze layer are depicted in Fig. 4. The training and internal validation loss are 0.2778±5.70E-02 and 0.2927±5.01E-02, respectively. There is a loss gap of around 5.37% between training and validation. The loss curve for validation shows that the best accuracy can be achieved at an epoch of 35. The time required for running the training with this method is 55.66±0.48s. The model's performance, measured by the intersection over union (IoU), for the modified U-Net architecture without a frozen layer is 0.5633±7.15E-02 during training and 0.5591±5.69E-02 during validation. The IoU difference between the training and validation results is approximately 0.74%.

The confusion matrix in Fig. 5 shows the performance of the modified U-Net architecture without the frozen layer, based on sampling 10 images. As depicted in Fig. 5(a), the model maintains stability in all stages, except image sample number 3, where a deviation in accuracy is observed (Fig. 6). The average accuracy during the training, validation, and

testing stages was 0.9951±2.92E-03, 0.9929±8.08E-03, and 0.9865±2.96E-02, respectively. The ANOVA test results showed no significant differences (p>0.05) in accuracy across these stages, implying that the model is steady and does not exhibit underfitting or overfitting.

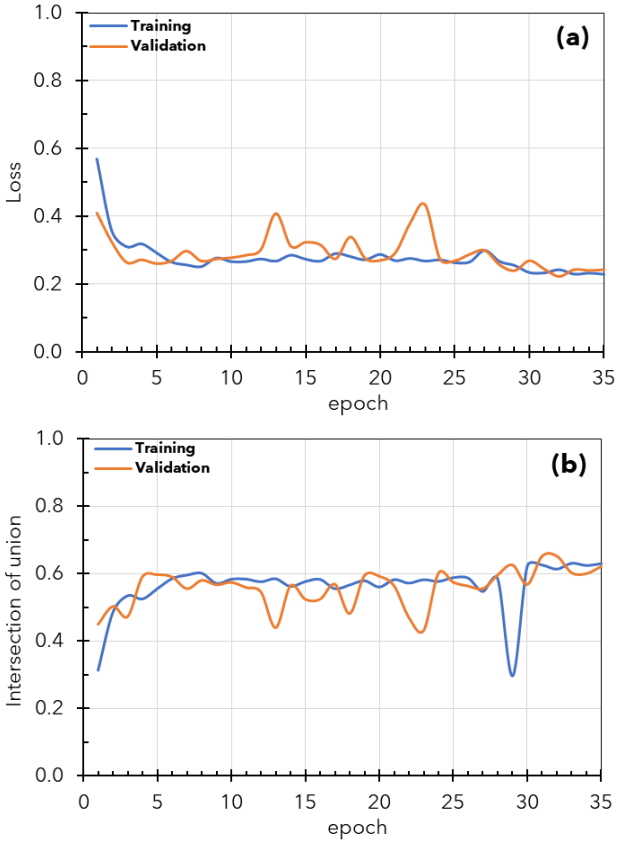


Fig. 4. Training and validation history of 35 epochs iteration from modified U-Net architecture without freeze layer model (a) accuracy and (b) IoU.

Precision as a key parameter for evaluating the model's consistency in predicting positive outcomes, yielded average scores of 0.7883±7.77E-02, 0.7511±2.04E-01, and 0.7592±2.18E-01 during the training, validation, and testing stages, respectively (Fig. 5(b)). Similarly, recall that measures the model's sensitivity in correctly predicting actual positive cases showed average values of 0.9784±5.25E-02, 0.8883±1.94E-01, and 0.8018±2.58E-01 across the same stages (Fig. 5(c)). Finally, the f1-Score, which balances precision and recall, averaged 0.8695±4.12E-02, 0.8029±1.82E-01, and 0.7584±2.21E-01 (Fig. 5(d)). An ANOVA analysis of these metrics (Precision, recall, and f1-Score) revealed no significant differences between the training, validation, and testing stages (p>0.05), indicating no false or incorrect patterns were identified in the dataset.

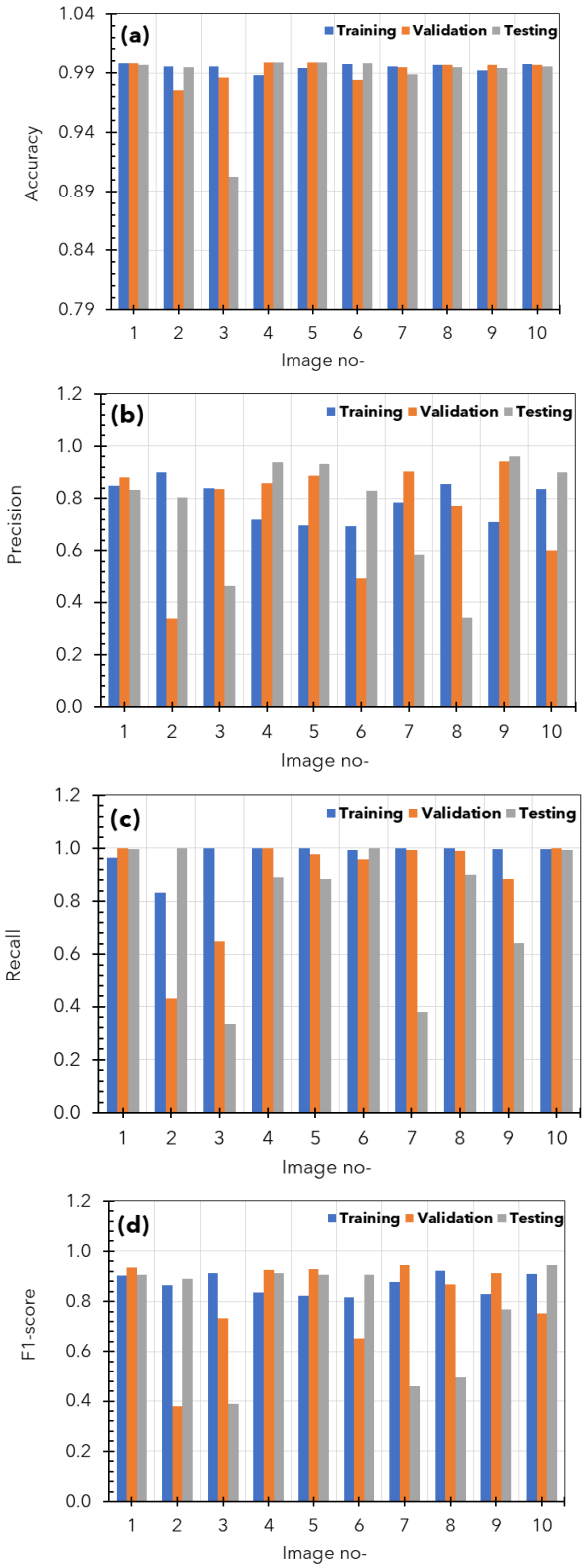


Fig. 5. Performance of sampling image from modified U-Net architecture without freeze in a confusion matrix (a) accuracy, (b) precision, (c) recall and (d) f1-score.

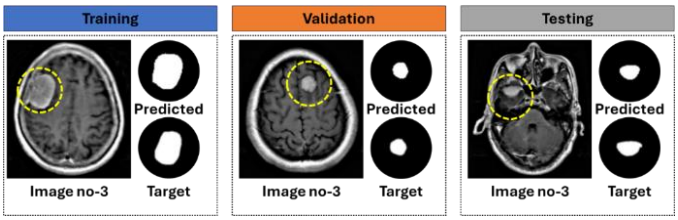


Fig. 6. Details of sample images 3 in training, validation, and testing by modified U-Net architecture without freeze model.

C. Modified U-Net Architecture with Freeze Layer

Fig. 7(a) illustrates the training history over 35 epochs using the modified U-Net architecture with a frozen layer, showing a training loss of 0.1947±7.71E-02 and a validation loss of 0.2468±3.86E-02. The loss difference between training and validation is approximately 26.73%. The training runtime for this method is calculated at 36.03±0.17s. Fig. 7(b) presents the model's performance, with an intersection over union (IoU) of 0.6764±8.10E-02 for training and 0.5995±5.14E-02 for validation, resulting in a 11.37% IoU difference between training and validation.

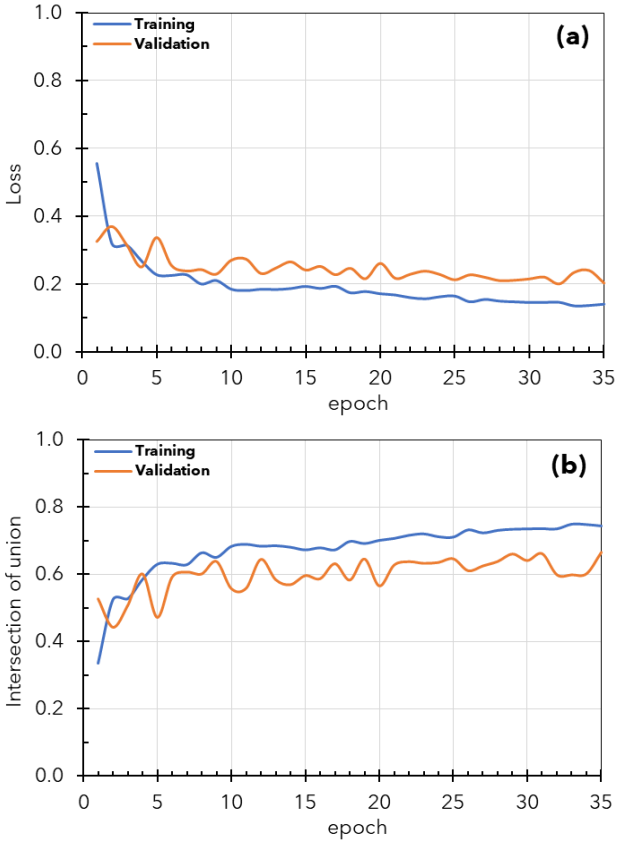
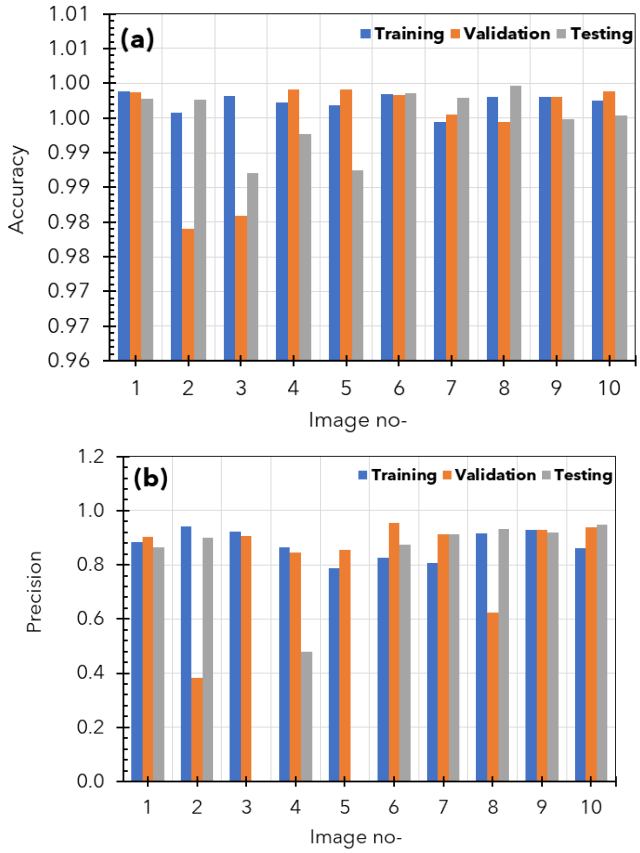


Fig. 7. Training and validation history of 35 epochs iteration from modified U-Net architecture with freeze layer model (a) accuracy and (b) IoU.

The confusion matrix in Fig. 8 illustrates the performance of the modified U-Net architecture with the frozen layer based on a sample of 10 images. As illustrated in Figure 8a, the model demonstrates stability across all stages, except image sample numbers 2 and 3, where a decline in validation accuracy is observed, and image sample number 5 shows a decrease in precision during the testing stage (Fig. 9). The average accuracy during the training, validation, and testing stages was 0.9973±1.35E-03, 0.9942±7.68E-03, and 0.9949±4.50E-03, respectively. ANOVA test results indicated that the model is stable and does not exhibit indications of underfitting or overfitting, as there were no significant differences in accuracy between these stages (p>0.05).

During the training, validation, and testing stages, the average precision scores for predicting favorable outcomes were 0.8742±5.43E-02, 0.8251±1.82E-01, and 0.6831±3.85E-01, respectively (Fig. 8(b)). Conversely, the recall metric, which quantifies the model's ability to accurately anticipate real positive instances, had mean values of 0.9549±6.05E-02, 0.8348±2.66E-01, and 0.8146±3.03E-01 over the same stages (Fig. 8(c)). In Fig. 8(d), the f1-Score, which measures the balance between accuracy and recall, had average values of 0.9109±3.92E-02, 0.8120±2.10E-01, and 0.7022±3.83E-01. Statistical study of the metrics (precision, recall, and f1-Score) using ANOVA showed no statistically significant variations throughout the training, validation, and testing stages (p>0.05). This suggests no erroneous or inaccurate patterns were detected in the dataset.



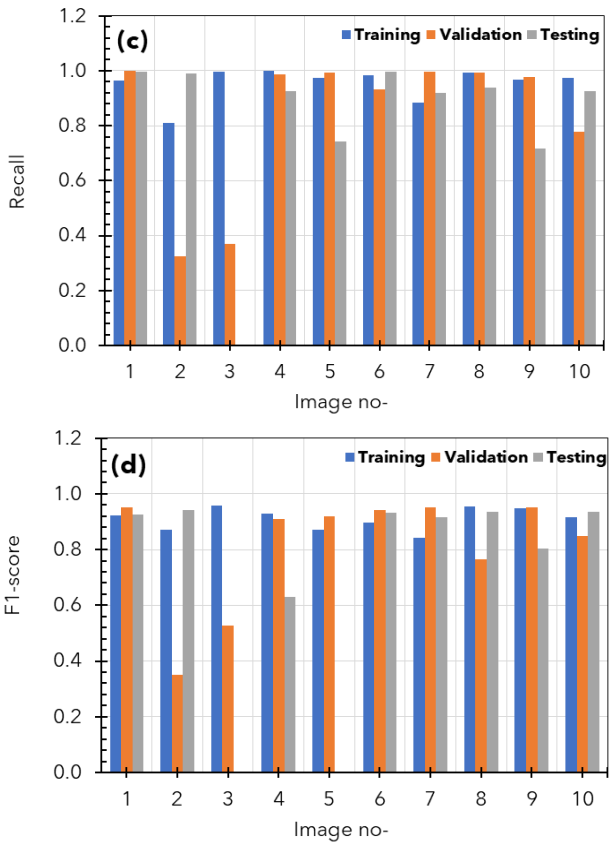


Fig. 8. Performance of sampling image from modified U-Net architecture with freeze in a confusion matrix (a) accuracy, (b) precision, (c) recall and (d) f1-score.

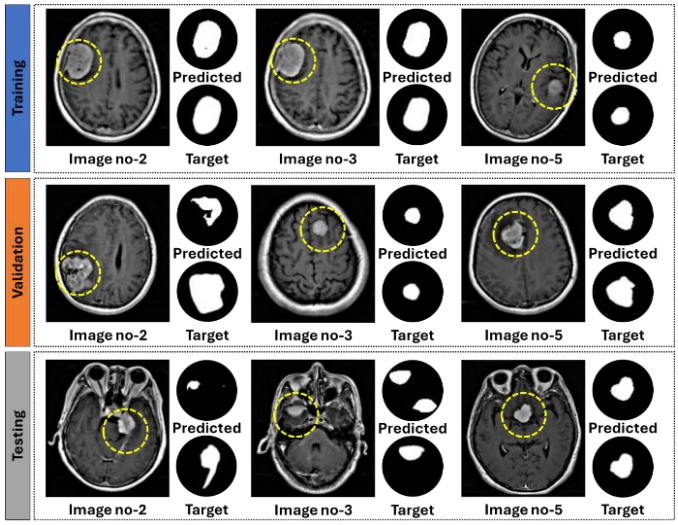
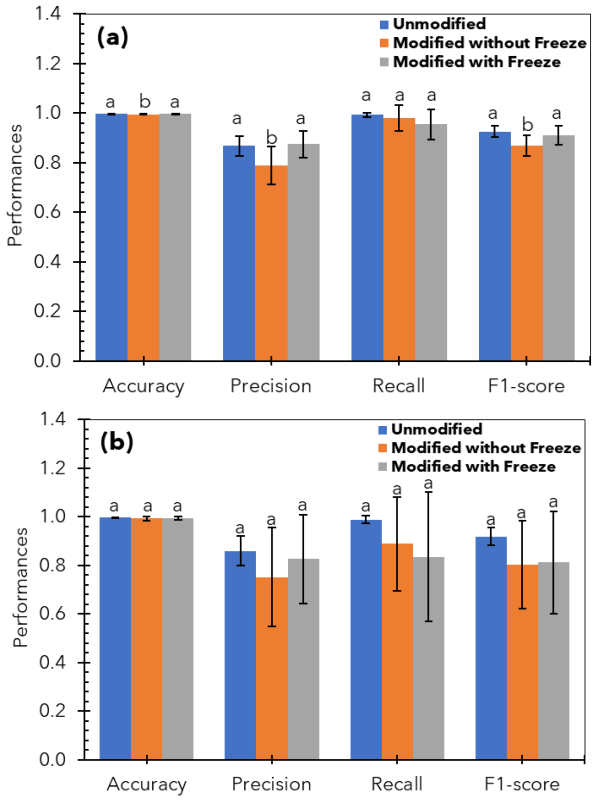


Fig. 9. Details of sample images 3 in training, validation, and testing by modified U-Net architecture with freeze model.

D. Comparison Performance Between the Proposed Method

The performance comparison of the three methods used in this study at each modeling stage is presented in Fig. 10. It can be seen that at the training stage for all parameters, there is a significant difference (p<0.05) between the three methods tested except for the recall parameter. However, no difference in performance was found between the three methods for the validation and testing stages (p>0.05). Finally, the method with modified U-net architecture with a freeze layer provides a significant difference in training time calculation. Meanwhile, the model performance is equally suitable for all three using unmodified U-Net architecture. Therefore, the method proposed through this study that combines modified U-net architecture with a freeze layer provides a more efficient calculation time efficiency of around 83% than unmodified U-net architecture. Compared with previous research, as reported by Cheng, et al. [26], this study's results have been better than those of prior research, as presented in Fig. 10(e).

The practical motivation for applying the theoretical results obtained in this study lies in improving the accuracy and efficiency of brain tumor segmentation in clinical settings. Accurate segmentation is crucial for diagnosing, planning treatment, and monitoring tumor progression, yet manual segmentation is time-consuming and prone to variability. Our modified U-Net architecture, enhanced with batch normalization, dropout, and freeze layers, offers a robust and scalable solution that addresses common issues such as overfitting and computational inefficiency, mainly when working with small MRI datasets. These theoretical improvements translate into more reliable, faster, and precise automated segmentation, ultimately aiding clinicians in making more informed decisions. The ability to generalize across different cases ensures that the model can be applied in real-world medical environments, improving diagnostic workflows and potentially leading to better patient outcomes through more personalized and accurate treatment strategies.



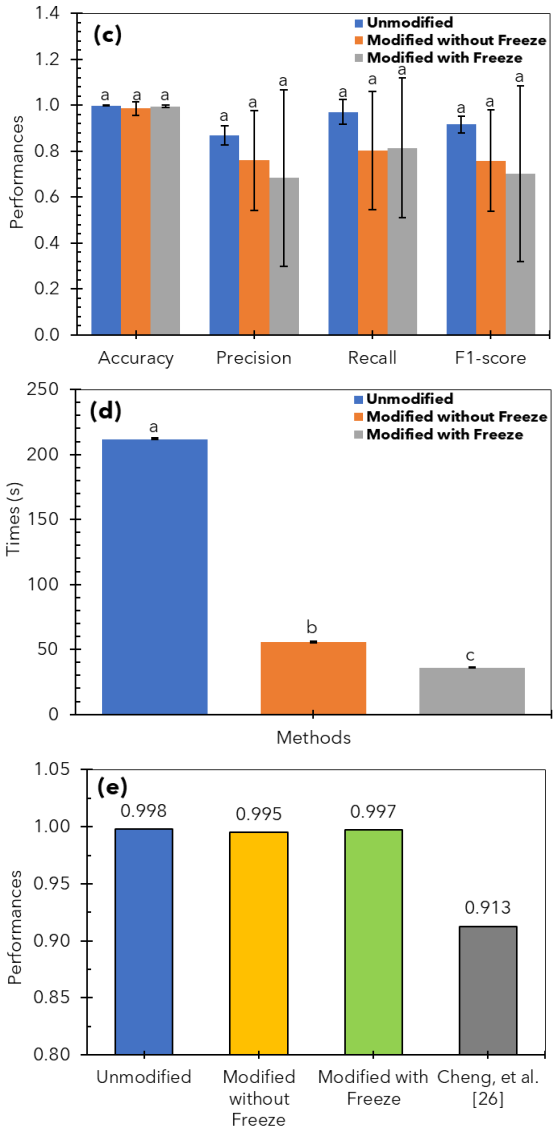


Fig. 10. Comparison performance between the proposed method on (a) training, (b) validation, (c) testing, (d) times and (e) comparison with a result from [26].

V. Conclusions and Future Work

In this study, we compared the performance of unmodified and modified U-Net architectures for brain tumor segmentation from MRI images, focusing on enhancements through batch normalization, dropout, and freeze layers. While the unmodified U-Net provided reliable performance, it exhibited a slight loss gap between training and validation, hinting at potential overfitting and resulting in lower intersection over union (IoU) scores. On the other hand, the modified U-Net without a freeze layer demonstrated improved convergence and reduced loss, with higher IoU scores indicating better segmentation capabilities. However, some inconsistencies were observed in precision and recall during the testing stage, showing that further optimization was necessary while the modifications improved performance. The most promising results came from the modified U-Net with a freeze layer, which achieved the highest IoU scores and maintained stability across all performance metrics, including accuracy, precision, recall, and f1-score. This model effectively mitigated overfitting by limiting the training depth while reducing the computational time required, making it highly efficient for practical use. The freeze layer modification proved particularly beneficial, allowing for a balance between model complexity and computational efficiency.

The modified U-Net architecture with a freeze layer significantly reduced training time by 81.72% compared to the unmodified U-Net while maintaining similar performance across validation and testing stages. Despite the differences observed in training, all three methods performed equally well regarding segmentation accuracy and consistency. While the unmodified U-Net is a solid baseline for segmentation tasks, the modified U-Net, especially with the freeze layer, shows superior performance, making it more suitable for real-world applications where accuracy and efficiency are paramount. Future studies could investigate further improvements to these architectures by incorporating advanced optimization methods and expanding the dataset to include more diverse tumor types and imaging conditions, thereby enhancing the model's generalizability.

References

[1] N. Weiskopf, L. J. Edwards, G. Helms, S. Mohammadi, E. Kirilina. Quantitative magnetic resonance imaging of brain anatomy and in vivo histology. Nature Reviews Physics. 2021. 3(8): 570-588.

[2] M. Martucci, R. Russo, F. Schimperna, G. D'Apolito, M. Panfili, A. Grimaldi, A. Perna, A. M. Ferranti, G. Varcasia, C. Giordano. Magnetic resonance imaging of primary adult brain tumors: state of the art and future perspectives. Biomedicines. 2023. 11(2): 364.

[3] F. Sanvito, A. Castellano, A. Falini. Advancements in neuroimaging to unravel biological and molecular features of brain tumors. Cancers. 2021. 13(3): 424.

[4] H. Saleem, A. R. Shahid, B. Raza. Visual interpretability in 3D brain tumor segmentation network. Computers in Biology and Medicine. 2021. 133: 104410.

[5] Y. Song, S. Ren, Y. Lu, X. Fu, K. K. Wong. Deep learning-based automatic segmentation of images in cardiac radiography: a promising challenge. Computer Methods and Programs in Biomedicine. 2022. 220: 106821.

[6] N. Yamanakkanavar, J. Y. Choi, B. Lee. MRI segmentation and classification of human brain using deep learning for diagnosis of Alzheimer's disease: a survey. Sensors. 2020. 20(11): 3243.

[7] M. Mittal, M. Arora, T. Pandey, L. M. Goyal. Image segmentation using deep learning techniques in medical images. Advancement of machine intelligence in interactive medical image analysis. 2020. 41-63.

[8] N. S. Punn, S. Agarwal. Modality specific U-Net variants for biomedical image segmentation: a survey. Artificial Intelligence Review. 2022. 55(7): 5845-5889.

[9] R. Azad, E. K. Aghdam, A. Rauland, Y. Jia, A. H. Avval, A. Bozorgpour, S. Karimijafarbigloo, J. P. Cohen, E. Adeli, D. Merhof. Medical image segmentation review: The success of u-net. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2024.

[10] R. Yousef, S. Khan, G. Gupta, T. Siddiqui, B. M. Albahlal, S. A. Alajlan, M. A. Haq. U-Net-based models towards optimal MR brain image segmentation. Diagnostics. 2023. 13(9): 1624.

[11] S. Das, M. k. Swain, G. K. Nayak, S. Saxena, S. Satpathy. Effect of learning parameters on the performance of U-Net Model in segmentation of Brain tumor. Multimedia tools and applications. 2022. 81(24): 34717-34735.

[12] N. Talat, A. Alsadoon, P. Prasad, A. Dawoud, T. A. Rashid, S. Haddad. A novel enhanced normalization technique for a mandible bones

segmentation using deep learning: batch normalization with the dropout. Multimedia Tools and Applications. 2023. 82(4): 6147-6166.

[13] M. U. Rehman, S. Cho, J. H. Kim, K. T. Chong. Bu-net: Brain tumor segmentation using modified u-net architecture. Electronics. 2020. 9(12): 2203.

[14] B. Lengerich, E. P. Xing, R. Caruana. On dropout, overfitting, and interaction effects in deep neural networks. arXiv preprint arXiv:2007.00823. 2020. 2.

[15] R. K. Samala, H.-P. Chan, L. M. Hadjiiski, M. A. Helvie, C. D. Richter. Generalization error analysis for deep convolutional neural network with transfer learning in breast cancer diagnosis. Physics in Medicine & Biology. 2020. 65(10): 105002.

[16] M. Iman, H. R. Arabnia, K. Rasheed. A review of deep transfer learning and recent advancements. Technologies. 2023. 11(2): 40.

[17] P. Jyothi, A. R. Singh. Deep learning models and traditional automated techniques for brain tumor segmentation in MRI: a review. Artificial intelligence review. 2023. 56(4): 2923-2969.

[18] T. A. Soomro, L. Zheng, A. J. Afifi, A. Ali, S. Soomro, M. Yin, J. Gao. Image segmentation for MR brain tumor detection using machine learning: a review. IEEE Reviews in Biomedical Engineering. 2022. 16: 70-90.

[19] J. Ramasamy, R. Doshi, K. K. Hiran, "Segmentation of brain tumor using deep learning methods: a review," in Proceedings of the International Conference on Data Science, Machine Learning and Artificial Intelligence, 2021, pp. 209-215.

[20] O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, 2015: Springer, pp. 234-241.

[21] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, K. H. Maier-Hein, "No New-Net," in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Cham, 2019: Springer International Publishing, pp. 234-244.

[22] S. Ioffe. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167. 2015.

[23] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research. 2014. 15(1): 1929-1958.

[24] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, B. Glocker. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical image analysis. 2017. 36: 61-78.

[25] K. Kamnitsas, W. Bai, E. Ferrante, S. McDonagh, M. Sinclair, N. Pawlowski, M. Rajchl, M. Lee, B. Kainz, D. Rueckert, "Ensembles of multiple models and architectures for robust brain tumour segmentation," in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers 3, 2018: Springer, pp. 450-462.

[26] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang, Q. Feng. Enhanced performance of brain tumor classification via tumor region augmentation and partition. PloS one. 2015. 10(10): e0140381.

[27] J. Cheng, W. Yang, M. Huang, W. Huang, J. Jiang, Y. Zhou, R. Yang, J. Zhao, Y. Feng, Q. Feng. Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation. PloS one. 2016. 11(6): e0157112.

Indoor Landscape Design and Environmental Adaptability Analysis Based on Improved Fuzzy Control

Jinming Liu[1], Qian Hu[2]*, Pichai Sodbhiban[3]
Modern Design Institute, Hunan Software Vocational and Technical University, Xiangtan, 411100, China[1, 2]
College of Art, Bangkok Thonburi University, Bangkok, 10170, Thailand[3]

***Abstract*—With the increasing demand for automation and intelligence in indoor landscape design, exploring efficient and precise control strategies has become particularly important. Robot-assisted technology and A* algorithm are utilized for indoor environment localization and mapping. Then, type-2 fuzzy adaptive fuzzy control is applied for indoor landscape automatic design. An improved genetic algorithm is utilized for environmental analysis to enhance the adaptability of indoor landscape design to the environment. In the results, the robot adopting this algorithm was significantly better than ordinary robots in path planning optimization, with a fitting accuracy of over 95%. The type-2 fuzzy control model had a maximum speed of 0.75m/s and an overshoot of only 7.1% for balancing robots, resulting in a faster recovery speed and smaller overshoot. The proposed method performed the best in terms of functionality, aesthetics, technicality, accessibility, and user satisfaction for landscape design effectiveness and environmental adaptability. The research improves indoor landscape design's automation. Meanwhile, the combination of fuzzy control and genetic algorithms enhances the design accuracy and environmental adaptability. This provides a new technological path for indoor landscape design.**

***Keywords*—Fuzzy control; indoor landscape design; environment; adaptability analysis; robot assisted**

I. Introduction

Indoor Landscape Design (ILD), as an important component of the built environment, is crucial for improving spatial quality and meeting functional requirements. However, when faced with the growing demand for personalization and diversity, traditional ILD shows significant limitations in terms of efficiency, accuracy, and adaptability. How to achieve efficient, precise, and environmentally adaptable design has become an urgent problem to be solved, especially in complex and ever-changing indoor environments [1]. As technology advances, automation and intelligence become new trends in design. Intelligent design not only improves design efficiency but also provides more accurate and personalized design solutions through data analysis and simulation [2-3]. Fuzzy Control (FC), as a control strategy that can handle uncertainty and fuzzy information, has shown its unique advantages in multiple fields. FC can effectively process fuzzy, inaccurate, or incomplete information by simulating human decision-making processes, thereby achieving optimized control in complex systems. Therefore, how to combine FC with ILD to improve the automation level and environmental adaptability of design has become a topic worthy of research [4-5]. To address this issue, this study proposes an improved FC-based ILD and environmental adaptability analysis method. This method combines robotics technology, A* algorithm, Range Type-2 Fuzzy Logic Control Mechanism (RT2FLCM), and Improved Genetic Algorithm (IGA). It aims to achieve automation and precision of ILD while improving environmental adaptability.

The innovation of the research lies in the application of fuzzy control theory to interior landscape design, and through the combination of robot-assisted technology and A* algorithm, the accurate location and mapping of indoor environment are realized. In addition, the study adopts type-2 fuzzy adaptive fuzzy control for interior landscape automatic design. The application of this method makes the design process more accurate and personalized. The contribution of the research is reflected in environmental analysis through improved genetic algorithms, which are capable of simulating natural selection and genetic mechanisms for global search and optimization. The application of this algorithm improves the accuracy and environmental adaptability of the design. By combining advanced control theory, algorithm and robot technology, a new analysis method of interior landscape design and environmental adaptability is proposed in this study, which provides a new technical path for the field of interior landscape design and has important theoretical and practical significance.

The article is divided into six sections. Section I is the introduction, through the background and research status of interior landscape design leads to the research theme. Section II is a literature review, which discusses and analyzes the research status of fuzzy control algorithm and interior landscape design at home and abroad. In Section III, robot assisted technology and A* algorithm are used for interior landscape positioning and environment mapping, and RT2FLCM is used for interior landscape automatic design. In Section IV, the effectiveness of the algorithm is verified by experiments. Section V is discussion, which discusses and analyzes the research results and compares them with other studies. Section VI summarizes the research results.

II. Related Work

When exploring the potential application of fuzzy logic in ecosystem service assessment, Biber et al. developed a

biodiversity-based fuzzy logic evaluation system. This evaluation system was applied to three different forest management scenarios. A simulation study was conducted for up to 100 years. These results confirmed the effectiveness of fuzzy logic as an evaluation tool [6]. Colak et al. proposed an active power factor correction method based on FC theory to address the growing demand for reliable and efficient power systems. Fuzzy logic controllers provided a highly adaptable and flexible solution by addressing the inherent uncertainty in power systems. In comparison with traditional control methods, the FC-based method had significant advantages in accuracy, robustness, and response time [7]. Hussein et al. proposed a fuzzy logic-based model for evaluating the spatial spaciousness. The proposed fuzzy model could accurately reflect input variables' influence on spatial spaciousness [8]. Khafajeh et al. started designing and developing an FC system for hydroponic greenhouses. Through optimization of the FC system, the average temperature during the day and night decreased from 34.25℃ and 23.22℃ to 31.17℃ and 21.96℃, respectively [9].

For intelligent drone control, Al Gizi A J H utilized a remote FC vehicle-mounted sonar tracking and detection device mounted to collect data for preventive maintenance of high-voltage power lines. Combining deep neural networks and FC achieved more efficient power line maintenance work [10]. Kasruddin et al. proposed a novel hybrid strategy combining spiral dynamic algorithm and other methods for flexible robotic arm systems' wheel hub angle tracking. This strategy not only accelerated convergence speed but also improved the solution accuracy. The optimized controller accurately tracked the expected response [11]. Incekara developed a fuzzy logic-based design method for primary school ergonomics classroom furniture. These results confirmed this fuzzy mathematical model's effectiveness [12]. Obinna proposed an intelligent spectrum allocation method to address the rapid increase in IoT devices and limited spectrum resources. This method utilized fuzzy logic to handle uncertainty and imprecise data. The method based on fuzzy logic effectively balanced channel availability and interference level, significantly improving service quality satisfaction [13].

To sum up, the application of fuzzy control is still in its infancy, especially in robot-assisted design and environmental adaptability analysis. A comprehensive method of interior landscape design and environment adaptability analysis is proposed by combining robot assistance technology, A* algorithm, type-2 fuzzy adaptive fuzzy control and improved genetic algorithm. This approach excelled in automation and precision, especially when dealing with uncertainty and complexity in the design process, showing superior performance over existing solutions. Compared with other studies, such as the fuzzy logic evaluation system based on biodiversity developed by Biber et al., the research method transforms the theory into practice through actual robot operation and automated design process, and improves the practical operability of the design. Compared with the active power factor correction method based on fuzzy control theory proposed by Colak et al., the application field of the research is more focused on interior landscape design, more targeted, and more in-depth in environmental adaptability analysis.

III. Methods and Materials

Firstly, the study utilizes robot assisted technology and A* algorithm for indoor landscape localization and environmental mapping. This helps robots to accurately draw and measure in complex indoor spaces. Then, RT2FLCM is utilized for automatic indoor landscape design. Finally, IGA is utilized for indoor environment analysis.

A. Robot Indoor Environment Mapping Based on A* Algorithm

In the automatic ILD, robot technology is adopted to assist in indoor spaces' construction and layout. The RB08 robot, as a compact and highly maneuverable multifunctional robot, is very suitable for complex operations in indoor environments. This robot has six degrees of freedom. Three degrees of freedom are specifically used for terminal positioning, while the rest are utilized to ensure precise positioning of the terminal [14-15]. SolidWorks software is utilized to construct robot's various joints. These joints are then imported into 3D Max to achieve coordinate transformation and relationship establishment between joints. Through this method, passive control between each joint can be achieved. This can precisely control the robot's movements, ensuring the automation and precision of ILD. By combining robotics technology and 3D modeling tools, ILD can not only improve efficiency, but also achieve more complex and refined designs. Fig. 1 shows the robotic arm's coordinate system in landscape design.

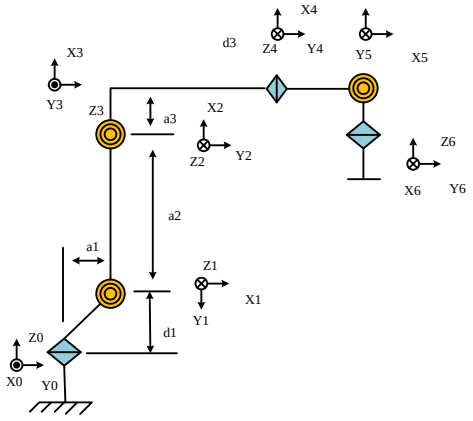


Fig. 1. Robot arm's coordinate system.

In Fig. 1, the D-H matrix is utilized to import the D-H parameter information into the connecting rod transformation matrix. The transformation matrices of each connecting rod are obtained, represented by Eq. (1).

$$A_i^{i-1} = A_{rot}(z_{i-1},\theta_i)A_{tran}(z_{i-1},d_i)A_{tran}(x_{i-1},\alpha_i)A_{rot}(x_{i-1},\alpha_i) \quad (1)$$

In Eq. (1), A_i^{i-1} refers to each connecting rod's transformation matrix. θ_i refers to the joint rotation angle. d_i refers to the joint displacement. α_i refers to the connecting rod's torsion angle. Based on two adjacent coordinate systems, using a homogeneous transformation matrix, the transformation matrix A_m^0 between the final attachment coordinate system and the base system of the m-freedom serial manipulator is obtained, represented by Eq.

(2).

$$A_m^0 = A_1^0 A_2^1 \cdots A_{m-1}^{m-2} A_m^{m-1} \tag{2}$$

In Eq. (2), A_m^0 represents the base system's transformation matrix. Any point's posture on the robotic arm's end connector can be described using the basic coordinate system, represented by Eq. (3).

$$p^0 = A_1^0 A_2^1 \cdots A_{m-1}^{m-2} A_m^{m-1} p^m \tag{3}$$

In Eq. (3), p represents any point's posture on the robotic arm's end connector. By multiplying the six joints' connection transformation matrices order, the terminal transformation matrix A_6^0 corresponding to the base coordinate system can be obtained, represented by Eq. (4).

$$A_6^0 = \begin{pmatrix} m_x & o_x & a_x & p_x \\ m_y & o_y & a_y & p_y \\ m_z & o_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4}$$

On this basis, a reverse kinematics method is proposed and transformed into reverse kinematics to calculate the corresponding joint parameters. The reverse calculation of motion is represented by Eq. (5).

$$\theta = [\theta_1 \theta_2 ... \theta_{m-1} \theta_m] = IKP(A_m^0) \tag{5}$$

In Eq. (5), IKP represents reverse motion. The indoor drawing path optimization of robots utilizes the A* algorithm, which is a widely utilized search algorithm in path planning. The A* algorithm excels at finding the shortest path from the starting point to the endpoint in complex graphical structures. The A* algorithm's core lies in its clever combination of heuristic functions and cost functions. Heuristic functions are utilized to predict the possible cost from the current node to the target node. The cost function evaluates the actual cost from the starting point to the current node [16-17]. Fig. 2 shows the A* algorithm.

In Fig. 2, the initial point is included in the OPEN list and the points in this table are checked. If the OPEN list is empty, the search is terminated and the path is reported as non-existent. If the OPEN list is not empty, the node n with the lowest $F(n)$ value is selected. Node n is removed from the OPEN list and added to the Completed list. Applying this algorithm to ILD can assist robots in precise drawing and measurement in complex indoor spaces. Designers can obtain higher quality spatial data, providing a solid foundation for creative design.

B. Indoor Landscape Design Based on Fuzzy Control

A step-by-step design method is adopted for a balancing robot's straight system. Firstly, a set of fuzzy logic adjusters is developed to address the characteristics and motion features of the balancing robot. Furthermore, by utilizing the membership functions of fuzzy sets, a preliminary architecture of an advanced fuzzy logic control system is constructed. Finally, detailed adjustments and optimizations are made to the overall equipment parameters to improve control effectiveness [18-19]. FC's core process is fuzzification, which is achieved by defining fuzzy sets and their membership functions. Within the determined input range U, usually within the interval of [0, 1], the fuzzy generator utilizes a specific mapping function to convert the exact value into a fuzzy value, represented by Eq. (6).

$$U \to [0,1], u \to \mu_A(u) \tag{6}$$

In Eq. (6), μ_A represents the uncertainty set A's membership function value. A fuzzer's core function is to convert the precise input data captured by the control mechanism into a set of uncertainties. Meanwhile, the membership function values are utilized to measure each component's membership strength. The strategy proposed by Zadeh for processing discrete information has been widely adopted. It expresses the fuzziness of input data through membership functions, enabling the control system to more effectively handle uncertainty and fuzziness, represented by Eq. (7).

$$A = \frac{A(u_1)}{u_1} + \frac{A(u_2)}{u_2} + \cdots + \frac{A(u_n)}{u_n} \tag{7}$$

In Eq. (7), u_n represents an element. $A(u_n)$ means membership degree. Type-2 fuzzy sets are extensions of Type-1 fuzzy sets that introduce additional dimensions to represent uncertainty. Assuming x is an element defined on a certain domain, its membership degree in a type-1 fuzzy set is represented as u. Therefore, the type-2 fuzzy set further refines x's membership degree by introducing an additional membership function, represented by Eq. (8).

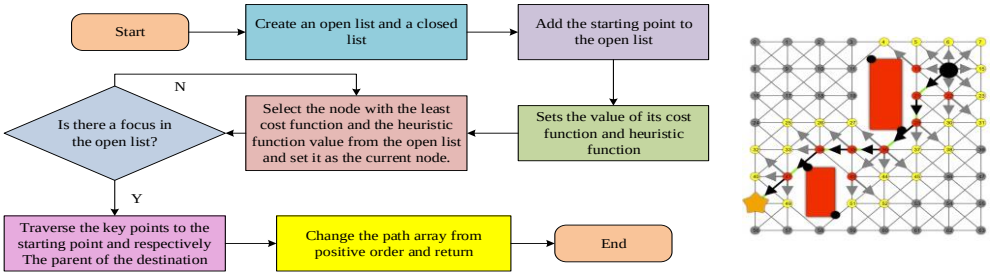


Fig. 2. Process of A* algorithm.

$$\tilde{A} = \left\{((x,u), u_{\tilde{A}}(x,u) \mid \forall x \in X, \forall u \in J_x \subseteq [0,1]\right\} \quad (8)$$

In Eq. (8), J_x represents the primary membership degree of x. $u_{\tilde{A}}$ means the sub-membership degree. Type-2 fuzzy set is an extension of Type-1 fuzzy set, which represents the uncertainty of membership degree by adding a dimension. The membership degree range of each type-1 fuzzy set is also from 0 to 1. However, when they combine, they form a more complex structure. This structure can express higher levels of uncertainty, thus forming a type-2 fuzzy membership function. In the advanced fuzzy sets, the Uncertainty Scope (US) is utilized to represent the set of all principal membership functions within the domain [20]. US provides a method for quantifying and visualizing the uncertainty of fuzzy sets. US covers all possible membership values, allowing for more accurate description and analysis of the characteristics of fuzzy sets. US is represented by Eq. (9).

$$FOU(\tilde{A}) = \bigcup_{x \in X} J_x \quad (9)$$

In Eq. (9), $\bigcup$ means the union of all principal membership functions. The uncertainty domain can also be represented in a coordinate system in Fig. 3.

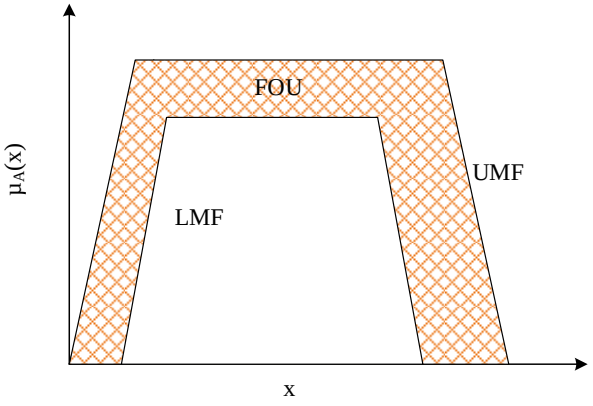


Fig. 3. Type-2 fuzzy set uncertainty field.

In Fig. 3, the boundary of the type-2 fuzzy set is determined by both the Upper Membership Function (UMF) and the Lower Membership Function (LMF). UMF reveals the highest possible membership degree that a set can achieve within the domain of discourse. LMF reveals the lowest membership degree. These two functions' difference forms an interval. This interval's shaded area represents the uncertainty domain, which refers to all possible membership degree of this set within these two boundaries. A system based on RT2FLCM is developed, which integrates the advantages of interval type-2 fuzzy sets to enhance adaptability and control effectiveness in uncertain and complex situations. Fig. 4 shows the structure of RT2FLCM.

In Fig. 4, RT2FLCM inherits the basic structure of type-1 FC, including three main steps: fuzzification, inference, and defuzzification. However, RT2FLCM introduces an additional component, a type reducer, after processing the rule base, which is utilized to convert type-2 fuzzy sets into type-1 fuzzy sets or specific values for further processing. This step is crucial to ensure that the system can smoothly execute subsequent control tasks. Through this design, RT2FLCM can more effectively manage and reduce uncertainty, improving control systems' performance and reliability.

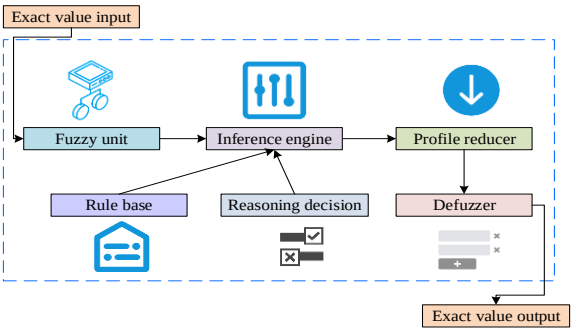


Fig. 4. The structure of RT2FLCM.

C. Indoor Environmental Analysis Based on Genetic Algorithm

For ILD and environmental analysis, an optimization technique that simulates natural selection and genetic mechanisms, namely GA, is utilized for environmental analysis and optimization. This method has significant parallel processing capabilities and global search advantages, enabling effective exploration in the parameter space in a probabilistic manner. It can automatically identify and guide the search process and dynamically adjust search strategies to adapt to constantly changing design requirements. Fig. 5 shows the GA.

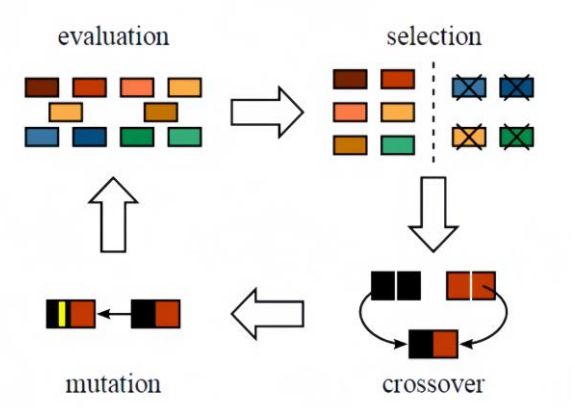


Fig. 5. Genetic algorithm flow chart.

In Fig. 5, the GA process includes encoding the initial population, using fitness values to measure the quality of chromosomes, and selecting individuals based on fitness values. Selection typically retains individuals with better fitness values. Crossover refers to gene exchange. Mutation refers to genetic modification. Adaptability is reflected in probability settings. When fitness begins to concentrate, the crossover and mutation probability increases to escape from local optima. When fitness is dispersed, reducing the crossover and mutation probability allows individuals to search for optimal solutions in their respective regions. The crossover probability is represented by Eq. (10).

$$P_c = \begin{cases} \dfrac{K_1(F' - F_{\min})}{F_{avg} - F_{\min}}, F' \le F_{avg} \\ K_2, F' > F_{avg} \end{cases} \quad (10)$$

In Eq. (10), P_c is the crossover probability. F is an individual's fitness utilized to perform mutation operations. F' means the smaller individual's fitness among two individuals that need to perform crossover operations. F_{avg} means the parental chromosome's mean fitness. $F_{\min}$ means the parent generation's minimum fitness. K is an adjustment parameter. The mutation probability is represented by Eq. (11).

$$P_m = \begin{cases} \dfrac{K_3(F - F_{\min})}{F_{avg} - F_{\min}}, F \le F_{avg} \\ K_4, F > F_{avg} \end{cases} \quad (11)$$

In Eq. (11), P_m is the crossover probability. When conducting ILD and indoor environment analysis, using GA alone requires multiple CFD simulations, which results in high computational costs. Neural networks are utilized as an alternative to CFD to reduce this cost, which are combined with GA to reduce the necessary computational workload. Fig. 6 shows the ILD process combining GA with neural networks and CFD.

In Fig. 6, during the initial stage of ILD, a neural network is utilized to predict the new design scheme's performance indicators. If these predicted results meet the expected design standards, CFD simulation is utilized to conduct the design scheme's in-depth analysis to obtain accurate CFD performance data. By combining neural networks and CFD, ILD can be more efficient and accurate, while providing designers with a powerful tool to achieve innovative and high-quality design results.

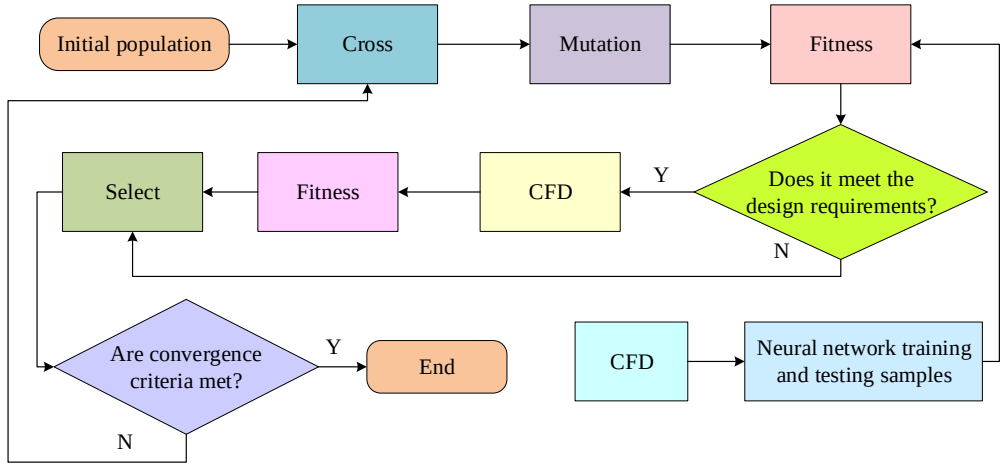


Fig. 6. Indoor landscape design process combining genetic algorithm with neural network and CFD.

IV. Results

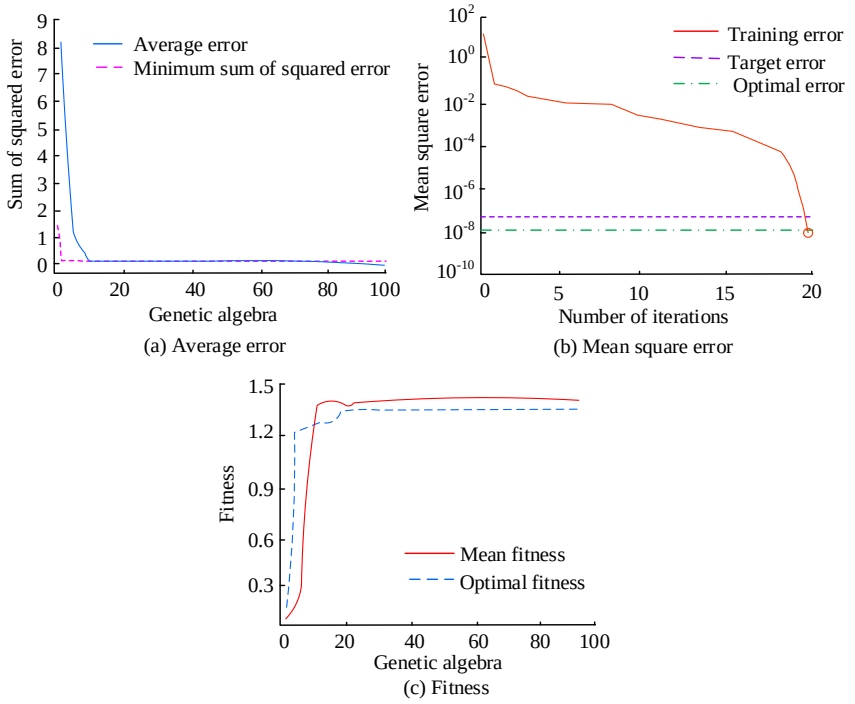
Firstly, the robots and algorithmic performance based on A*+IPID+IGA was evaluated. The mean square error, sum of squared errors, and fitness curve of A*+IPID+IGA were analyzed. Then, simulation analysis was conducted on the type-2 FC model. Finally, the landscape design effects and environmental adaptability of different methods were compared and analyzed.

A. Performance Evaluation of Robots and Algorithms based on A*+IPID+IGA

Simulation experiments were conducted on robots based on A*+IPID+IGA using Matlab2018b. Figure 7 shows the mean square error, Sum of Squared Error (SSE), average fitness, and optimal fitness curves of A*+IPID+IGA.

In Fig. 7 (a), the average and minimum SSE of A*+IPID+IGA converged after only 6 iterations, and SSE remained stable at 0.19. In Fig. 7 (b), A*+IPID+IGA showed a faster iteration speed in the first 7 iterations. However, from the 8th to the 19th iteration, the iteration speed slowed down slightly. Finally, at the 20th iteration, the algorithm achieved convergence, indicating that the algorithm improved the convergence rate during the iteration. In Fig. 7(c), A*+IPID+IGA achieved convergence between average and optimal fitness in the first 20 iterations. This method showed a fast convergence rate at the beginning of the iteration and reached the convergence point at the 20th iteration, with the final convergence value stabilizing at 1.44. The path planning optimization accuracy for indoor landscape mapping and design of robots based on A*+IPID+IGA was analyzed. The study compared the real and planned trajectories of ordinary robots and robots based on A*+IPID+IGA in Fig. 8.

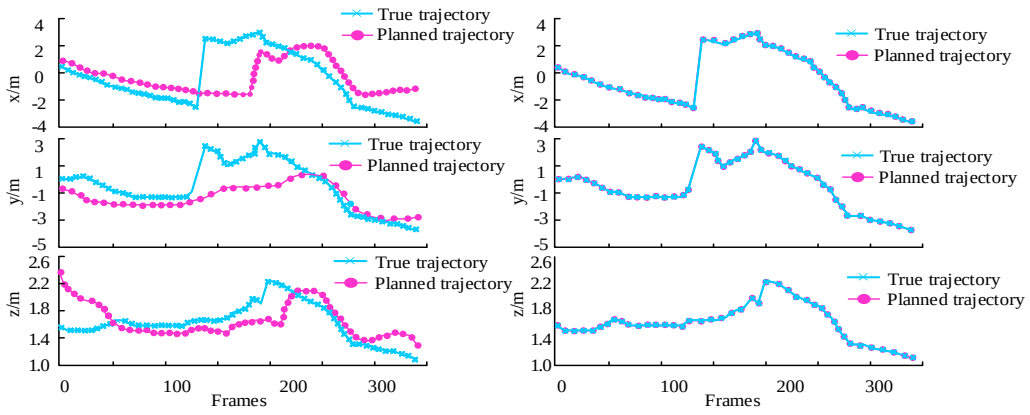
In Fig. 8 (a), the fitting accuracy of the robot's real and planned trajectories based on A*+IPID+IGA reached over 95%. In contrast, the fitting accuracy of ordinary robot path planning in Fig. 8 (b) was only about 67%. Therefore, robots based on A*+IPID+IGA had significant optimization improvements in path planning optimization compared to ordinary robots.



(a) Average error (b) Mean square error

(c) Fitness

Fig. 7. Mean square error, SSE, and optimal fitness curve.



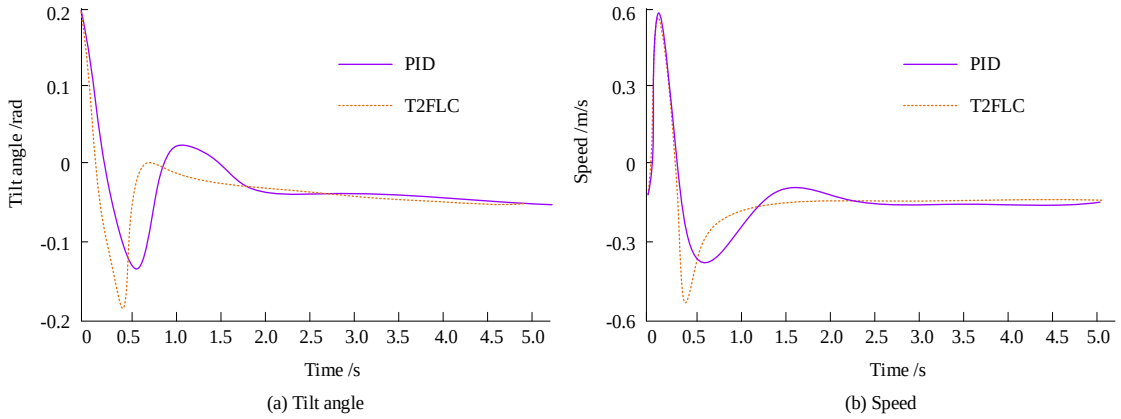
(a) Ordinary robot path planning (b) Robot path planning based on A*+IPID+IGA

Fig. 8. Comparison of paths and trajectories of robots.

B. Simulation Analysis of Robot Type-2 Fuzzy Control

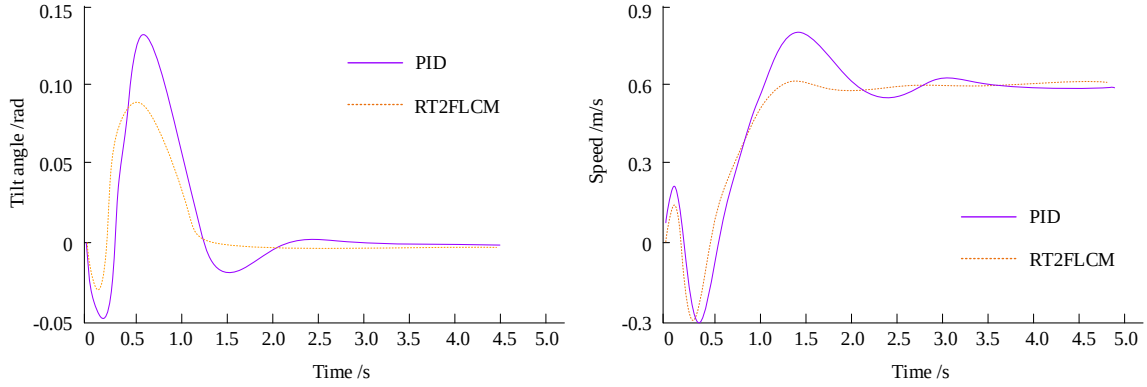
The research assumed that robot's intrinsic parameters were not only accurate and error-free, but also remained constant during operation. External interference factors were excluded. The initial tilt angle was set to 0.3 radians, with the target velocity set to zero, to evaluate the robot's ability to restore balance to a specific tilt angle. The experiment utilized the Simulink module of MATLAB, set the target speed to zero, and set the initial state array. The study compared the classical Proportional-Integral-Derivative (PID) regulation technique with the fusion function integrated RT2FLCM. Fig. 9 shows the robot's tilt angle and velocity changes.

In Fig. 9 (a) and Fig. 9(b), when initially tilted at 0.3 radians, the robot increased to a speed of nearly 0.6m/s in a short period of approximately 0.2s. As the speed of the robot increased, the tilt rapidly decreased and exceeded the equilibrium point, entering a reverse tilt state, indicating that the robot was starting to tilt backwards. Subsequently, the robot began to slow down its speed and gradually returned to equilibrium after experiencing two oscillations. When using traditional PID control, the robot's maximum tilt angle reached 0.16 radians. In contrast, when using the RT2FLCM strategy, although the tilt angle was slightly larger, RT2FLCM significantly improved the recovery speed. Specifically, the robot restored the tilt angle and velocity to zero in about 1.7s. This meant the robot reached a balanced and stationary state, while PID took about 2.5s. After evaluating the stability adjustment ability, further testing will be conducted on robot performance in rate adjustment. This robot was accelerated from a stationary state to a target speed of 0.7m/s while maintaining its initial state unchanged. Fig. 10 shows the changes in speed and tilt angle under two control modes.



(a) Tilt angle (b) Speed

Fig. 9. Tilt and velocity response curve.



(a) Tilt angle (b) Speed

Fig. 10. Velocity and tilt curves under two control modes.

In Fig. 10(a), PID had a maximum negative tilt angle of 0.08, while RT2FLCM had a maximum negative tilt angle of 0.04. PID had a maximum positive tilt angle of 0.2, while RT2FLCM had a maximum positive tilt angle of 0.14. In Fig. 10(b), PID's maximum speed reached 0.9m/s, with an overshoot of 28.6%. In contrast, the maximum speed under RT2FLCM control was 0.75m/s, with an overshoot of only 7.1%. Using PID caused two oscillations, while RT2FLCM only produced one oscillation. In terms of reaching the target speed, the type-2 FC was completed within 1.6s, while using a dual closed-loop PID required approximately 3s. During the overall speed adjustment process, RT2FLCM exhibited smoother and faster performance, with better control effects.

C. Analysis of Landscape Design Effects and Environmental Adaptability

A comparative analysis was conducted between the proposed ILD (Method 1), ILD based on GA (Method 2), ILD based on linear programming (Method 3), ILD based on deep learning (Method 4), and ILD based on building information model (Method 5). The comparative indicators were normalized. Table I shows the final ILD effect.

TABLE I. COMPARISON OF INDOOR LANDSCAPE DESIGN EFFECTS

Method index	Method 1	Method 2	Method 3	Method 4	Method 5
Functionality	0.93	0.71	0.63	0.77	0.83
Aesthetic	0.97	0.68	0.75	0.71	0.85
Technicality	0.94	0.75	0.69	0.82	0.91
Environmental quality	0.89	0.73	0.81	0.79	0.88
Aesthetic value	0.85	0.79	0.75	0.68	0.74
Accessibility	0.92	0.82	0.84	0.65	0.82
Social benefit	0.83	0.81	0.82	0.63	0.78
Innovativeness	0.91	0.88	0.73	0.74	0.86
User satisfaction	0.95	0.74	0.78	0.77	0.90

In Table I, Method 1 performed the best in terms of functionality, aesthetic, technicality, accessibility, and user satisfaction, with scores of 0.93, 0.97, 0.94, 0.92, and 0.95, respectively. Method 5 also performed well on most indicators, especially in terms of functionality and user satisfaction, with scores of 0.83 and 0.90, respectively. In contrast, Methods 2, 3, and 4 scored lower on some indicators, indicating that their performance in these areas needed improvement. Overall, Method 1 became the most popular and effective design method due to its comprehensive advantages. Table II shows the environmental adaptability of five methods.

TABLE II. ENVIRONMENTAL ADAPTABILITY OF THE FIVE METHODS

0	Method 1	Method 2	Method 3	Method 4	Method 5
Functional adaptability	0.95	0.77	0.67	0.67	0.83
Human body engineering	0.91	0.73	0.59	0.82	0.88
Thermal comfort	0.86	0.82	0.68	0.73	0.91
Acoustic adaptability	0.88	0.79	0.81	0.76	0.83
Light environment	0.94	0.61	0.73	0.68	0.85
Air quality	0.92	0.64	0.75	0.85	0.88
Materials and finishes	0.91	0.81	0.80	0.65	0.79
Spatial flexibility	0.86	0.69	0.69	0.74	0.73
Psychological comfort level	0.88	0.78	0.74	0.71	0.68

In Table II, Method 1 performed outstandingly in terms of functional adaptability, human body engineering, thermal comfort, light environment, and air quality, with scores of 0.95, 0.91, 0.86, 0.94, and 0.92, respectively. This demonstrated its outstanding performance in meeting human needs and environmental comfort.

V. DISCUSSION

Improved fuzzy control and genetic algorithm are applied in interior landscape design to improve the automation and accuracy of the design and enhance the environmental adaptability. In terms of functionality, the proposed method shows a high score of 0.93, which is in contrast to the results obtained by Moreno et al. [1] in applying fuzzy logic to the preventive protection and restoration monitoring of heritage buildings. While Moreno et al.'s study focused on evaluation and monitoring, this study applies fuzzy control to the design process itself, achieving a higher functional score. In terms of aesthetics, the score of the study is 0.97, which is compared with the study of Hussein et al. [8] using fuzzy logic to evaluate the spaciousness of architectural design studios, which mainly focuses on the physical properties of space, while this study comprehensively considers aesthetics and provides a more comprehensive design scheme. In terms of environmental adaptability, the interior landscape design method proposed in this study got a score of 0.86 in terms of thermal comfort, which was compared with the result of Khafajeh et al. [9] applying fuzzy logic in the hydroponic greenhouse control system, which mainly focused on the optimization of environmental control, while this study took environmental adaptability as a part of the design process. To achieve a more comfortable indoor environment. Through these comparisons, it can be seen that the proposed method has obvious advantages in automation design and environmental adaptability. The research not only improves the accuracy of the design, but also significantly improves the automation level and environmental adaptability of the design through the combination of fuzzy control and genetic algorithm.

VI. CONCLUSION

A method for analyzing ILD and environmental adaptability based on improved FC is proposed. Its effectiveness is verified through experiments. In the results, A*+IPID+IGA showed good convergence performance during the iteration. The mean square error and SSE converged rapidly after 6 iterations and reached a stable state within 20 iterations. The average and optimal fitness also showed a rapid convergence trend. These results validated this algorithm's effectiveness and reliability in solving optimization problems. For path planning optimization, robots based on A*+IPID+IGA showed higher fitting accuracy compared to ordinary robots, reaching over 95%. The ordinary robots' fitting accuracy was only about 67%. These demonstrated the optimization capability of the improved algorithm in path planning. The ILD and environmental adaptability analysis method based on A*+IPID+IGA has demonstrated excellent performance in robot path planning, balance control, speed control, and landscape design effect evaluation. This method not only improves the automation and accuracy of design, but also significantly enhances environmental adaptability. However, there are still some shortcomings in the research. For example, it is assumed that robot's internal parameters are constant without considering external interference factors. This may affect the control effect in practical applications. Future research can further explore the impact of parameter changes and external disturbances on system performance, as well as how to optimize algorithms to adapt to more complex real-world environments.

REFERENCES

[1] Moreno M, Prieto A J, Ortiz R. Preventive conservation and restoration monitoring of heritage buildings based on fuzzy logic. International Journal of Architectural Heritage, 2023, 17(7): 1153-1170.

[2] Çalışkan B. Spatial Suitability Scoring For Rail Transit Network Extension by Developing 'Rail Transit Fuzzy Logic Model' (RFLM) in Matlab Environment. International Review for Spatial Planning and Sustainable Development, 2024, 12(1): 222-241.

[3] Ting L, Khan M, Sharma A, Ansari M D. A secure framework for IoT-based smart climate agriculture system: Toward blockchain and edge computing. Journal of Intelligent Systems, 2022, 31(1): 221-236.

[4] Zhu Q, Kumar P M, Alazab M. Computer application in game map path-finding based on fuzzy logic dynamic hierarchical ant colony algorithm. International Journal of Fuzzy Systems, 2022, 24(5): 2513-2524.

[5] Àgueda A, Vacca P, Planas E, Pastor E. Evaluating wildfire vulnerability of Mediterranean dwellings using fuzzy logic applied to expert judgement. International journal of wildland fire, 2023, 32(6): 1011-1029.

[6] Biber P, Schwaiger F, Poschenrieder W, Pretzsch H. A fuzzy logic-based approach for evaluating forest ecosystem service provision and biodiversity applied to a case study landscape in Southern Germany. European Journal of Forest Research, 2021, 140(6): 1559-1586.

[7] Colak A M. Parameters Optimized with the Pattern Search Algorithm for Fuzzy Control Based Active Power Factor Correction. Electric Power Components and Systems, 2024, 52(7): 1115-1128.

[8] Hussein A Y, Mustafa F A. A quality learning environment: an assessment of spaciousness of architectural design studios of Iraqi Kurdistan region universities using fuzzy logic. Open House International, 2024, 49(1): 181-204.

[9] Khafajeh H, Banakar A, Minaei S, Delavar M. A hydroponic greenhouse fuzzy control system: design, development and optimization using the genetic algorithm. Spanish Journal of Agricultural Research, 2023, 21(1): e0201-e0201.

[10] Al Gizi A J H. UAV Flight Fuzzy Controller with Deep Learning Network Fault Checker of High-Voltage Lines. International Journal of

Intelligent Systems and Applications in Engineering, 2023, 11(2): 877-891.

[11] Kasruddin Nasir A N, Ahmad M A, Tokhi M O. Hybrid spiral-bacterial foraging algorithm for a fuzzy control design of a flexible manipulator. Journal of Low Frequency Noise, Vibration and Active Control, 2022, 41(1): 340-358.

[12] Incekara C O. Post-COVID-19 ergonomic school furniture design under fuzzy logic. Work, 2021, 69(4): 1197-1208.

[13] Obinna O R, Joseph E C. DYNAMIC SPECTRUM ALLOCATION IN COGNITIVE RADIO NETWORKS FOR IoT DEVICES USING FUZZY LOGIC. Advance Journal of Science, Engineering and Technology, 2024, 9(3): 1-14.

[14] Zhang Y, Chen J, Wang Q, Tan C, Li Y, Sun X, Li Y. Geographic information system models with fuzzy logic for susceptibility maps of debris flow using multiple types of parameters: a case study in Pinggu District of Beijing, China. Natural Hazards and Earth System Sciences, 2022, 22(7): 2239-2255.

[15] Gupta N, Lee S H. Trapezoidal interval type-2 fuzzy analytical hierarchy process technique for biophilic element/design selection in lodging industry. Journal of the Operational Research Society, 2023, 74(7): 1613-1627.

[16] Chiang T Y. Refurbishment criteria performance assessment methodologies based on a multiple-criteria approach. Journal of Housing and the Built Environment, 2021, 36(1): 263-282.

[17] Jiang Y, Fakhruldeen H, Pizzuto G. Autonomous biomimetic solid dispensing using a dual-arm robotic manipulator. Digital Discovery, 2023, 2(6): 1733-1744.

[18] Xiong Y, Pan L, Xiao M, Xiao H. Motion control and path optimization of intelligent AUV using fuzzy adaptive PID and improved genetic algorithm. Mathematical Biosciences and Engineering, 2023, 20(5): 9208-9245.

[19] Surendra H J, Deka P C. Municipal residential water consumption estimation techniques using traditional and soft computing approach: a review. Water Conservation Science and Engineering, 2022, 7(1): 77-85.

[20] Choudhuri S, Adeniye S, Sen A. Distribution Alignment Using Complement Entropy Objective and Adaptive Consensus-Based Label Refinement For Partial Domain Adaptation. Artificial Intelligence and Applications, 2023, 1(1): 43-51.

Optimising Delivery Routes Under Real-World Constraints: A Comparative Study of Ant Colony, Particle Swarm and Genetic Algorithms

Rneem I. Aldoraibi, Fatimah Alanazi, Haya Alaskar, Abed Alanazi
Department of Computer Science-College of Computer Engineering and Science in Al Kharj,
Prince Sattam bin Abdulaziz University P.O. Box 151, Alkharj 11942, Saudi Arabia

***Abstract*—Effective logistics systems are essential for fast and economical package delivery, especially in urban areas. The intricate and ever-changing nature of urban logistics makes traditional methods insufficient. Hence, requirements for the application of sophisticated optimisation techniques have increased. To optimise package delivery routes, this study compares the performance of three popular evolutionary algorithms: ant colony optimisation (ACO), particle swarm Optimisation (PSO), and genetic algorithms (GA). Finding the best algorithm to minimise delivery time and cost while taking into account real-world limitations, such as delivery priority. This guarantees that deliveries with a higher priority are prioritised over others, which may substantially impact route optimisation. We examine each algorithm to create the best possible route plans for delivery trucks using actual data. Several factors are employed to assess each algorithm's performance, including robustness to changes in environmental variables and computational efficiency—the simulation models delivery demands using actual data. Results indicate that ACO performed better in Los Angeles and Chicago, completing the shortest routes with respective distances of 126,254.18 and 59,214.68, indicating a high degree of flexibility in intricate urban layouts. With the best distance of 48,403.1 in New York, on the other hand, GA achieve good results, demonstrating its usefulness in crowded urban settings. These results highlight how incorporating evolutionary algorithms into urban logistics can improve sustainability and efficiency.**

Keywords*—Evolutionary algorithms; genetic algorithm; particle swarm optimisation; ant colony optimisation; urban logistics; route optimisation

I. Introduction

Over the last few decades, global greenhouse gas emissions have increased considerably, which are widely considered as the primary contributors of climate cshange [1]. Meanwhile, there has been a growing consciousness of the environmental impacts of business activities, encouraging companies, researchers and governments to discover optimal solutions that support operations with sustainability principles within logistics. Logistics lead to a large dimension of greenhouse gases, making it essential to adopt greener and more sustainable logistics systems [2], [3], [4]. Such efforts not only benefit the planet but also cater to an increasingly eco-conscious human population. Sustainability has increasingly become a concern among academics and practitioners in the fields of logistics and supply chain management. Although research in this area has progressively increased, it still presents multiple directions worthy of exploration [3], [5], [6]. In particular, there is a crucial need for investigations that illuminate how businesses approach sustainability and how organisations and researchers can develop more sustainable supply chains. Enhancing the sustainability of supply chains is considered a fundamental strategic action towards achieving sustainable development goals. Logistics and supply chain performances account for at least one-third of energy consumption and one-third of GHG emissions [7]. To moderate these effects, it is essential to adopt a systemic approach to transformative change in our supply chains—from production to distribution—to enhance their sustainability. It is essential to consider the effects of supply chains and logistics on sustainability from social and economic perspectives. This includes issues with equity, labour conditions and employability. Logistics was first used in the military to describe the methods used by soldiers to acquire, store and transport supplies and equipment [1]. Managing the movement of resources throughout the supply chain is the focus of logistics, a term that is now frequently used in the corporate world, particularly in manufacturing [8]. This covers gathering, storing and moving resources to their intended locations. In the context of logistics, sustainable transport refers to the application of procedures and tools that lessen the impact of transportation and distribution operations on the environment. This idea goes beyond just transporting products from one place to another; it takes into account the environmental impact of these activities [9]. The goals of sustainable logistics are to maximise resource utilisation, minimise greenhouse gas emissions and lessen air and water pollution [5]. Furthermore, businesses can minimise the number of trucks required for deliveries by optimising delivery routes, which can save fuel expenses and maintenance costs and can enhance resource efficiency. In addition, it will enhance customer satisfaction. Companies can boost customer satisfaction and can encourage repeat business and customer loyalty by promptly and efficiently providing goods and services.

Many different optimisation techniques have been developed based on computational intelligence, such as evolutionary algorithms and solutions that opened up the domain of metaheuristics. This study compares the performance of three optimisation algorithms to determine which is better for scheduling problems in logistics: ant colony optimisation (ACO) [10], particle swarm optimisation (PSO) and genetic algorithms (GAs) [11]. The ACO algorithm is a probabilistic optimisation approach inspired by nature that

finds the most effective paths in complex surroundings by simulating the foraging behaviour of ants. GAs are adaptive heuristic search algorithms that iteratively develop a population of solutions to address optimisation and search problems. They are based on the concepts of natural selection and genetics. PSO is another computational approach. It was motivated by the social behavior of fish schools and flocks of birds.

The paper aims to examine three models for logistic delivery optimisation using ACO, PSO and GA, addressing the particular difficulties presented by urban delivery environments. It aims to address a routing problem, where the objective is to determine the shortest path to deliver packages while accounting for the importance of each delivery. Priorities have an impact on the cost estimate. Thus, it is essential to manage more essential deliveries early in the route in addition to reducing the overall journey distance. In this paper, Los Angeles, New York and Chicago are used as study locations for package delivery optimisation research. By choosing these cities, we can address a wide range of variables and complexities that are reflective of worldwide urban logistics difficulties. These cities have a lot of business activity and e-commerce transactions, making them essential economic hubs. As a result, there are several delivery operations for both arriving and departing goods. Gaining knowledge from these cities will help in designing scalable, reliable and effective delivery methods that can be applied to different metropolitan environments globally. It will enhance the research's application value in real-world scenarios.

This study could add significantly to the topic of optimizing urban logistics in a number of ways. First, it offers an analysis of three popular algorithms (ACO, PSO, and GA) in relation to actual urban delivery problems, such as delivery priority. Second, the study assesses algorithm performance using a real data set, guaranteeing applicability and practical relevance. Third, the study provides useful insights for choosing the best method based on the particular needs of urban logistics by analyzing computational efficiency and route optimization accuracy.

The paper is structured as follows, the second section highlights previous research on the application of ACO, PSO, and GA in logistics and other relevant fields. The third section will introduce the methodology which we describe in the experimental design, including the setup of the optimization algorithms. This section also details the criteria for performance evaluation, focusing on delivery efficiency. The result section presents the findings from our simulations, comparing the effectiveness of ACO, PSO, and GA in optimizing delivery routes under urban constraints. The discussion section presents our findings for urban logistics systems. Finally, the conclusion section that addressed the key findings and discussed the broader implications for urban logistics optimization.

II. Literature Review

In this section, existing approaches and previous studies will be studied. Al-Tayar and Alisa [6] proposed several scenarios to obtain the optimal routing path in stochastic networks, including working according to static and dynamic network data. They used the evolutionary algorithm ACO to discover the optimal routing path between the source node and the target node witch's helpful to obtain optimisation that increases logistical effectiveness that leads to contributing to environmental sustainability by minimising fuel consumption, reducing emissions and conserving resources. However, there is a need to work on dynamic data in real time that can be modelled using other probability distributions.

Popović et al. [12] suggest to increase the efficiency of its logistics operations by improving the methodology for evaluating logistics processes using a new model. This model involves the creation of a novel grey full-consistency method which is used to calculate the weight values of the strengths, weaknesses, opportunities and threats factors of a logistics company However, because the application of the approach in the logistics field had not been considered previously, there are some limitations in developing the approach, such as the complex mathematical process for computing criteria weights even if it is applicable in evaluation processes in other various fields.

Zhai [13] proposed addressing the green low-carbon logistics path optimisation problem using the snowmelt heuristic optimisation algorithm. It starts by analysing the characteristics of the green low-carbon logistics path optimisation problem, then considers the optimisation cost and conditional constraints of the green low-carbon logistics path optimisation problem and uses the snowmelt heuristic algorithm model afterwards. The author compared the results with those of many different algorithms and discovered that the snowmelt heuristic algorithm achieves better performance overall but can easily fall into the local optimum problem.

Franco et al. [14] proposed a system that was designed by machine learning algorithms. To provide a solid and sustainable solution to make route adjustments such as re-routing and re-scheduling of the delivery for unpredicted cases, the possibility of a difference between the planned and the actual delivery routes, which is why the use of technology to respond efficiently to all possible events may happen. Even their conceptual framework still needs to be applied in different situations, adapted and extended, and at the same time, it helps in finding a common ground to feed data and obtain values, especially with a new tendency these days to synchronise digital technologies' penetration with all aspects of life.

Sadeghi and Haapala [15] expanded the research on previous work by incorporating the carbon cost into the mathematical cost to improve the mathematical model. The biomass-to-bio-oil supply chain (BTBSCS) used mobile and stationary bio-refineries and contributed to the literature on bio-oil problems by designing a genetic algorithm to obtain a near-optimal solution for establishing mobile and stationary bio-refineries to reduce logistical and carbon costs. Researchers must simultaneously improve the economic, environmental and social performance of bioenergy supply chains. Gao, Cao [16] focused on redesigning a new sustainable reverse logistics supply chain network with the existing forward logistics supply chain network by taking into account economic, environmental and social sustainability. Then, they proposed the MOSINP model to formulate the problem of sustainable reverse logistics supply chain network redesign. Their goal is to support logistics

activities in the face of uncertain demand for new products and the return volume of products of multiple quality levels.

Weber et al. [17] proposed different optimisation problems to find suitable, qualified and optimal solutions for sustainability. The results of this study indicate that there are many improvement models that address the three dimensions of sustainability simultaneously and the social dimension of sustainability is the least studied aspect. An applied classification of mathematical modelling approaches used in sustainable societies is provided. Another research should shift the focus from models that deal primarily with economic and environmental aspects to more balanced models that include all three aspects of sustainability. This paper promotes the transparent and rapid communication of research that highlights the role of optimisation in interdisciplinary fields of mathematical programming and provides SI optimisation models with relevant sustainability indicators.

Zarbakhshnia et al. [18] suggested a probabilistic mixed-integer linear programming model for a sustainable forward and reverse logistics network problem that takes into account many products, stages, periods, and objectives. Their mode aimed to find a new environmental constraint and social matters in the objective functions as its innovation and contribution. This model is based on using a non-dominated sorting genetic algorithm. Their result achieved a better performance compared with a multi-objective PSO.

III. METHODOLOGY

ACO, PSO and GAs will be constrained in this paper. Metrics, including best route, distance, time computation consumption and environmental impact, will be the main topics of comparison. The goal was to find the best set of routes for vehicles that minimise the overall trip distance and cost given a set of delivery destinations, a depot and restrictions on vehicle capacity and delivery times. Every package in our delivery system has a priority feature that is dynamically determined by how likely it is to be delivered late. The 'late delivery risk' feature is a function that prioritises packages to reduce delays and improve customer satisfaction. The main objective is to guarantee on-time delivery by modifying operating priorities according to projected delivery dates and times.

A. Ant Colony Optimisation (ACO)

The ACO technique was inspired by the foraging behaviour of ant colonies, first introduced by Dorigo [10]. Ants are eusocial insects that rely on a community-based approach for survival, rather than existing as individual species. They communicate with each other through sound, touch and pheromones. Pheromones are chemical compounds secreted by ants that trigger social responses within the same species. These chemicals act similarly to hormones but are external to the body and affect the behaviour of other ants. Because most ants live on the ground, they also communicate by leaving pheromone trails on the soil surfaces, which can be detected and followed by other ants [19].

The ACO algorithm, a probabilistic optimisation strategy inspired by nature that finds the most efficient paths in complex surroundings by mimicking the foraging behaviour of ants, is used to tackle this task.

Because ants live in a community of nests, the fundamental idea of ACO is to track how ants leave their nests to get food by taking the shortest route possible. Initially, ants start randomly moving in the space around their nests to search for food. That randomised search technique opens up multiple possible routes from the nest to the food source. Ants now bring some of the food back with them, increasing pheromones along their route, depending on the kind and amount of food they discover.

Depending on these pheromone trials, the probability of selection of a specific path that has been performed by following the ants' path would be a guiding factor to the food source. Evidently, this probability is based on the rate of concentration and evaporation of pheromones. It can also be observed that when the evaporation rate of pheromones is also one of many different deciding factors, the length of each path can easily be accounted for.

- Pheromone Model: A pheromone matrix was initialised to influence the probability of including each city in a route.
- Construction of Solutions: Ants constructed solutions by probabilistically choosing the next city to visit based on a rule combining pheromone strength and a heuristic function (distance and priority).
- Pheromone Update: After all ants completed their routes, pheromones were updated based on the quality of the solutions, with more successful routes receiving higher pheromone levels.
- Daemon Actions: Optional global updates were performed to intensify or diversify the search.
- Iteration: The search continued until a stopping criterion, such as the maximum number of iterations for convergence, was met.

In his experiments, the pheromone updating process is modified as well as the solution representation to prioritise specific deliveries in (ACO). The updating of pheromone trails according to the nodes' priorities. This adjustment makes sure that during the optimisation process, high-priority deliveries are given preference and precedence over others.

Pseudocode for ACO

Input:

Distance matrix, Priority, Number of ants, Number of iterations, Evaporation rate ρ, Alpha α, Beta β

Initialize:
Initialize pheromone levels on paths (tau)
Initialize heuristic information (eta), such as 1/distance
Initialize priority scores for nodes (priority)
For each iteration:
for each ant:
Place ant at a starting node
while ant has not completed its tour:
Select the next node based on transition

probabilities:
Move the ant to this next node
Optionally update pheromones on the path (local update)
Assess the quality of the ant's tour
Update pheromones based on the quality of the tour (global update)
Identify the best tour of the iteration based on lowest cost or highest priority fulfillment
Return the best overall tour found during all iterations

Therefore, the calculation takes into account the tour duration and a priority factor of this tour, which may depend on the nodes that the ant visited to update the pheromones. Routes that efficiently visit high-priority nodes can be prioritised by this factor.

B. Genetic Algorithm Optimisation (GA)

GAs are defined as adaptive heuristic search algorithms that belong to a larger part of evolutionary algorithms [11], [20]. They are predicated on concepts from genetics and natural selection. These are effective uses of the random searches made possible by past data to focus the search on the area of the solution space where performance is better. This is often used to produce solutions to search and optimization issues.

GAs have simulated the process of natural selection, which means that those species that can be able to change in the environment can survive, reproduce and go to the next generation [11]. In simply, that mimics the "survival of the fittest" by using individuals from succeeding generations to solve an issue. Every generation is made up of a population of individuals, and each individual is a potential solution as well as a point in search space. Every person has been represented as a bit string, integer, float, or string of characters. That string is analogous to the chromosome.

In this experiment, we modified the fitness evaluation to incorporate priorities. For instance, routes that visit high-priority locations earlier might receive higher fitness scores. The selection process will typically be implemented using methods such as roulette wheel selection, where the probability of an individual being selected is proportional to its fitness. This step ensures that higher-quality (higher fitness) individuals are more likely to be selected. The following section will describe the algorithm steps:

- Initialisation: The population was initialised with randomly generated possible routes. Each route or chromosome corresponds to a complete solution to the route.
- Fitness Function: The fitness of each chromosome was evaluated based on the total route distance, cost and adherence to delivery priority.
- Selection Process: A tournament selection process was used to select parent solutions for crossover based on their fitness scores.
- Crossover and Mutation: Ordered crossover (OX) and swap mutation were applied to generate new offspring, ensuring genetic diversity and exploration of the solution space.
- Termination: The algorithm terminated after a fixed number of generations or if there was no improvement in the best solution for a consecutive number of generations.

Pseudocode for GA

Input: Distance matrix, Priority, Population size N, Number of generations G, Crossover rate Pc, Mutation rate Pm, Fitness function F
Output: Best solution found
Initialize:
The population with N random solutions
Evaluate the fitness of each individual in the population using F
For each generation from 1 to G:
Select parents from the current population based on their fitness
Perform crossover on the parents to form new offspring, with probability Pc
Apply mutation to the offspring, with probability Pm
Evaluate the fitness of the new offspring using F
Select individuals for the next generation from the current population and the new offspring
If any offspring is better than the best solution found so far:
Update the best solution
Return the best solution

C. Particle Swarm Optimisation (PSO)

PSO has been inspired by a swarm of birds or a school of fish. At the same time, the algorithm is called a population-based stochastic algorithm, and it was developed by Russell et.al. in 1995. That is the overall concept of PSO and the basis of its biological phenomena [23]. In this paper, fitness should take priority into account in addition to the objective function (such as cost or distance), making sure that solutions that complete high-priority tasks are given a higher evaluation. The following section will explain the algorithm steps:

- Particle Representation: Each particle represents a potential solution to the route, encoded as a sequence of delivery points.
- Velocity and Position Update: Customised velocity and position update rules suitable for combinatorial problems were implemented, focusing on sequence operations, such as swaps, influenced by velocity vectors.
- Fitness Evaluation: Similar to GA, the fitness of each particle was assessed based on the route's total distance, cost and adherence to delivery priority.
- Global and Personal Best: Particles updated their velocities towards their personal best and the global best positions found during the search process.

- Convergence: The algorithm ran for a predetermined number of iterations or until performance plateaued.

Pseudocode for PSO

Input: Distance matrix, Priority, Number of particles N, Number of iterations I, Inertia weight W, Cognitive component C1, Social component C2, Objective function ObjFunc
Output: Best known position gBest
Initialize swarm of N particles with random positions and velocities
Evaluate the fitness of each particle using ObjFunc
Set pBest of each particle to its initial position
Set gBest to the position of the best performing particle in the initial swarm
For each iteration from 1 to I:
For each particle p in the swarm:
Update velocity and position of particle p:
Evaluate the fitness of the updated position of particle p
If the fitness of the updated position is better than the fitness at pBest:
Update pBest to the new position
If the fitness of the updated position is better than the fitness at gBest:
Update gBest to the new position
Return gBest

IV. Data Preparation

To verify the proposed sustainable logistics optimisation, an experiment has been designed and the experimental data are an extension of an open-source reference to address the logistics optimisation problem. In our experiment, the dataset used contains 25837 rows and 53 columns, with three cities being selected from that dataset (Chicago, New York and Los Angeles). Each city has several predefined delivery locations (latitude, longitude). Each location will have associated package delivery requirements, including package priority (late delivery). The priority information feature, that is, each algorithm's decision-making, is influenced by the corresponding priorities assigned to each node. The probability of reaching that node sooner increases with higher priorities.

The dataset was prepared before starting use in this study using many different techniques to ensure that any blanks are removed. In addition, there are no duplications or symbols in the dataset. Table I shows the number of addresses in each city.

TABLE I. Data Set

City	*Number of address*
Chicago	3885
Los Angeles	3417
New York	1816

This experiment adapts PSO to a complex permutation issue by reinterpreting the particle movement in the solution space, effectively using swaps influenced by pseudo-velocity. An element of complexity common to scheduling and logistics operations in the real world is added by the adaption, which involves managing priorities in route planning. This method exemplifies the adaptability of PSO and its potential for combinatorial optimisation beyond conventional uses.

Performance was evaluated based on the quality of the solution (total travel distance and cost), computational time and robustness against variations in problem parameters. A combination of the distance and strategic significance of the nodes visited may be used to determine which tour is optimal for each iteration, with an emphasis on routes that better meet higher-priority requirements.

V. Results and Evaluation

In our experimental setup, we implemented three well-known optimisation algorithms, namely, ACO, PSO and GA, to address a complex delivery routing problem and priority considerations. Each algorithm was configured with specific parameters tailored to balance evaluation and manipulation as shown in Table II, with ACO using 10 ants, PSO comprising 30 particles and GA operating with a population of 100 individuals. Our scenarios, which were generated from real-world data, involved more than 1000 delivery points with diverse priority levels. Each algorithm was executed 10 times per scenario to ensure robustness, with results focusing on efficiency, compliance with time computational consumption priority constraints and computational performance.

TABLE II. Parameters of Algorithms

Parameter	*GA*	*PSO*	*ACO*
Population Size	population = 100	num_particles = 30	number of ant =10
Operators	Ordered Crossover (OX), simple swap mutation Implements tournament selection	Velocity Update	Pheromone Update
Crossover Rate	0.7	-	-
Mutation Rate	0.05	-	-
Stopping Criteria	number of generation 100	number of iteration 100	number of iteration 100
Coefficients	-	alpha=1.0, beta=2.0, evaporation=0.5	Alpha (pheromone importance) =1 Beta (heuristic importance) =1

In Fig. 1 to Fig. 9, the figures visualize the optimization process of each evolutionary algorithm, where the Y-axis represents the best fitness value and the X-axis refers to the number of generations. This axis shows the number of generations through which the algorithm has been processed. As can be seen from Fig. 1 to Fig. 9, the performance of each algorithm improves with each iteration, as indicated by the decreasing score of the best distance. The performance of the GA algorithm improves consistently across all three experiments. In contrast, PSO shows that an increase in the number of iterations does not necessarily enhance its performance.

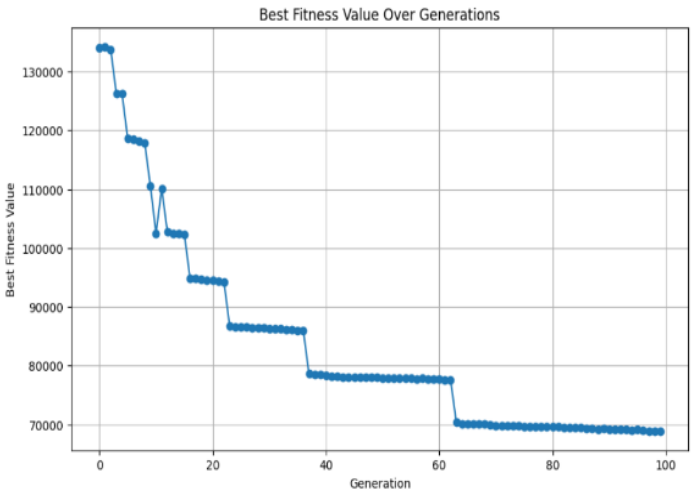


Fig. 1. Performance of GA for New York.

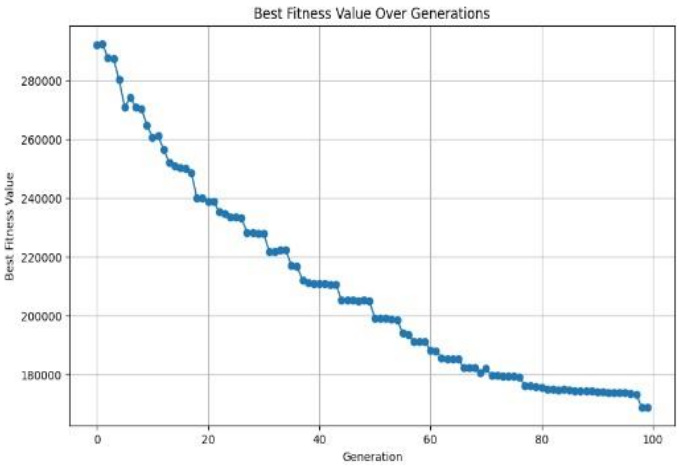


Fig. 2. Performance of GA for Chicago.

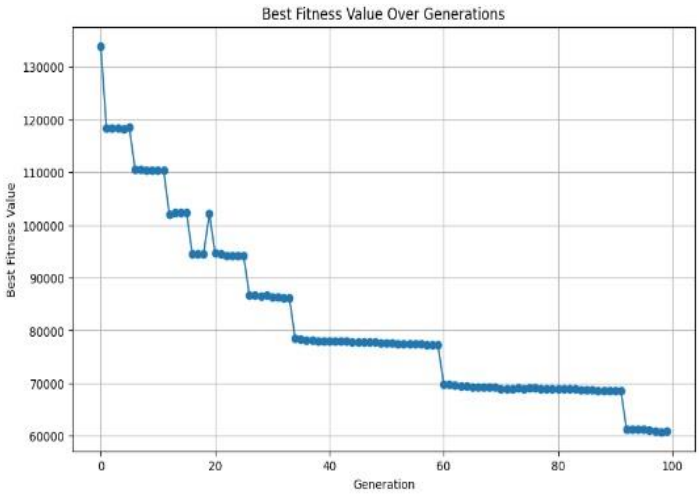


Fig. 3. Performance of GA for LA.

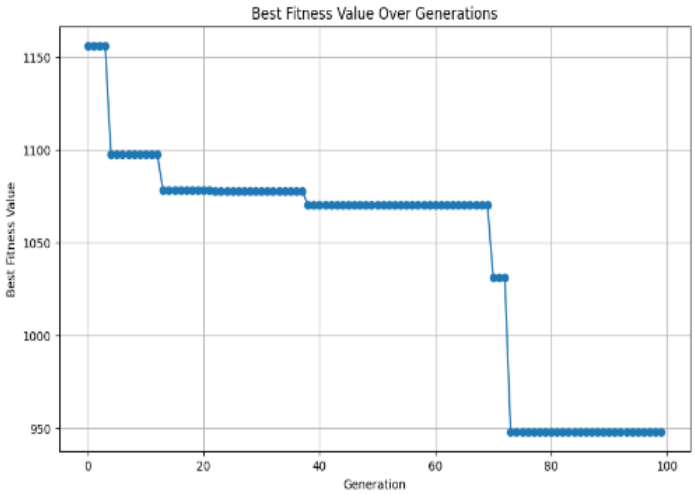


Fig. 4. ACO performance for LA.

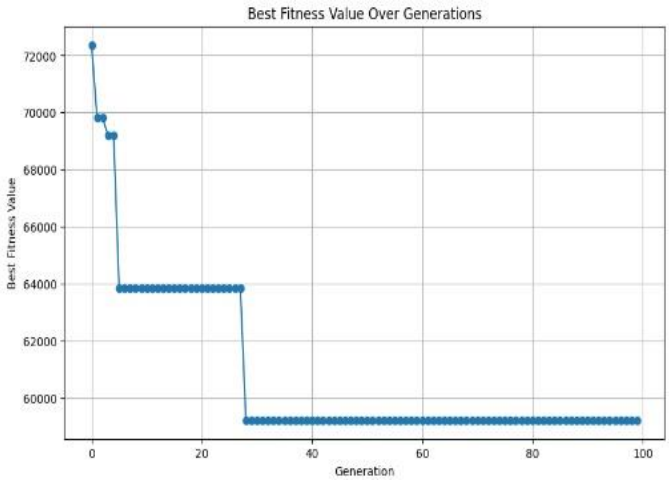


Fig. 5. ACO performance for Chicago.

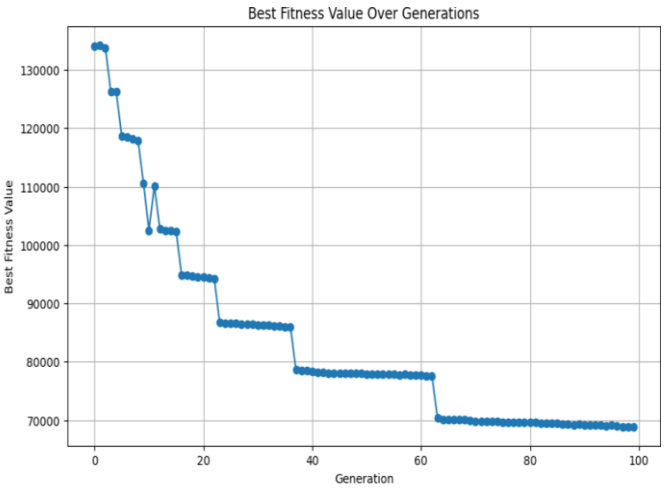


Fig. 6. ACO performance for New York.

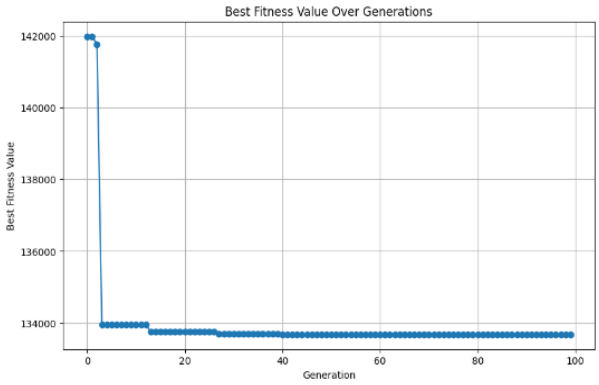


Fig. 7. PSO performance for LA.

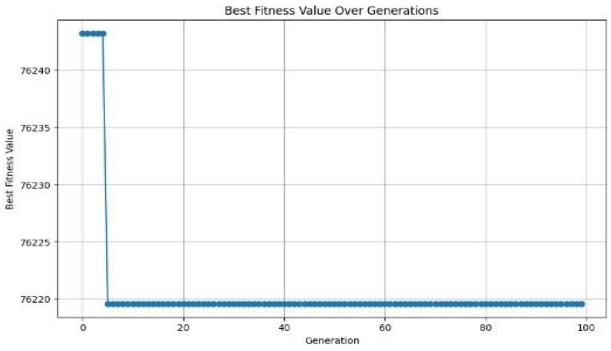


Fig. 8. PSO performance for New York.

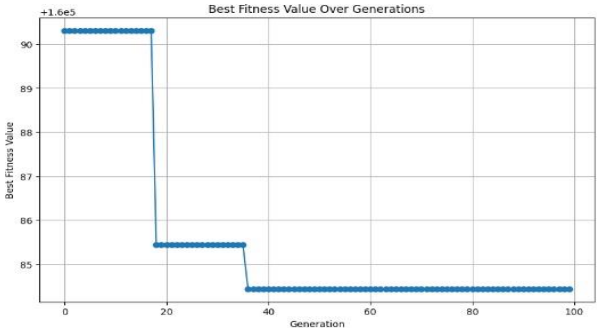


Fig. 9. PSO optimisation performance for Chicago.

Table III shows the result of each algorithm. The best performance was achieved by ACO for Chicago and LA, where the lowest performance was achieved in New York. However, when focused on time consumption, PSO is the fastest algorithm, and this is related to the lowest computational complexity in the algorithm procedure.

TABLE III. RESULT OF THE THREE OPTIMIZATION ALGORITHMS

Data set	Evaluation matrix					
	The best distance			*Time*		
	ACO	***GA***	***PSO***	***ACO***	***GA***	***PSO***
Chicago	59214.681	169017.6	160084.44	1156.282	285.6	27.92
LA	948.1877	60893.05	70260.64	560.4	130.49	20.59
New York	126254.18	48403.1	76219.58	8173.28	180	12.11

VI. DISCUSSION

Different performance characteristics for each algorithm under the urban package delivery scenario were revealed by our comparison investigation of GA, PSO and ACO. ACO had the best overall efficiency, the quickest delivery times and the lowest operating expenses. This was followed by GA, which demonstrated strong performance but was less efficient in terms of time and cost, and ACO, has achieved the best score based on time consumption.

The efficacy of ACO in our investigation is consistent with previous research that emphasizes its advantages in continuous optimisation issues, which we have converted to a discrete context by carefully adjusting parameters.

Given that ACO performed better in our simulation, its implementation may greatly improve operational effectiveness in urban logistics. Urban delivery services might undergo a revolution as a result of the shorter delivery times and cheaper operating costs, which would benefit consumers by lowering prices and increasing profitability for logistics companies. ACO's capacity to address constraints, such as the delivery priority problem, demonstrates its usefulness in scenarios where route dependability is crucial, reinforcing the suggestions made by [21] for logistics applications. In multi-objective logistical scenarios, GA demonstrated its versatility as observed by [22], as evidenced by its competitive performance to achieve a good result even with constraints.

This modified ACO algorithm works especially well in delivery and logistics settings where some deliveries are more essential than others possibly because of commodities that must be delivered on time. It enables the algorithm to automatically modify its pathfinding to give these crucial nodes a higher priority, mirroring operational priorities found in automated decision-making systems in the real world. Furthermore, ACO may be especially helpful in situations where late deliveries result in significant fines, such as in the case of medical or just-in-time industrial supplies, because of its high reliability in meeting delivery timetables.

Our results have several noteworthy implications, one of which is the possibility of minimising environmental effects through optimal routing. The effectiveness of ACO could contribute to decreased emissions and fuel consumption, which helps sustainability objectives in logistics and urban planning. This is essential as cities worldwide work to address climate change and lower their carbon footprints.

Although results show promise, they are restricted to simulated settings and might not accurately represent the intricacies of human dynamics and real-world traffic patterns. To evaluate and improve the models, future research should try to apply these algorithms in real-time logistics operations, possibly through pilot programmes.

VII. CONCLUSION

This research not only confirms that ACO, PSO and GA are appropriate for optimising urban logistics but also creates chances for integrating these algorithms into practical uses. Through extensive simulations and analysis, we coucnlued that these algorithms could significantly enhance operational

efficiency by optimizing delivery routes. Our data indicated that ACO, especially, excelled in handling complex urban environments by adapting to dynamic constraints. The integration of these algorithms into real-world data set shows promising potential to improve the operational success of logistics firms. By implementing these optimized strategies, companies can more reliably meet delivery schedules. The practical deployment of these methods led to a measurable improvement in punctuality and efficiency, as evidenced by the decrease in average delivery times and which will increase customer satisfaction rates.

Finally, this paper focused on the problem of package delivery route optimisation in heavily populated urban states with obtained results. We recommend modifying the ACO algorithm as a solution. Throughout the project, an effective model that addressed the problems with urban delivery was created, and using outside literature, we assessed the effectiveness of our approach using a variety of indicators. Although benchmark tests were ideal solutions, our findings were affected by algorithmic complexity and real-world application considerations. Overall, with space for improvement and parameter adjustment, our ACO-based approach provided insights into effective route planning for food delivery services in urban settings.

Because traffic congestion constantly develops in urban areas because of an increase in traffic vehicles, this project is advantageous for traffic routing in urban areas with more complex types of roads. This package delivery route optimisation project can be performed on a bigger scale, for example, by creating a routing system for an entire nation's traffic network. The data and performance measures that were covered in this project can help with upcoming studies and initiatives that might use them as a point of reference.

ACKNOWLEDGMENT

The authors are very grateful to thank their Deanship of Research and Graduate studies for technical and financial support in publishing this work successfully.

REFERENCES

[1] Mikhaylov, N. Moiseev, K. Aleshin, and T. Burkhardt, "Global climate change and greenhouse effect," Entrep. Sustain. Issues, vol. 7, no. 4, p. 2897, 2020.

[2] M. Massar, I. Reza, S. M. Rahman, S. M. H. Abdullah, A. Jamal, and F. S. Al-Ismail, "Impacts of autonomous vehicles on greenhouse gas emissions—positive or negative?," Int. J. Environ. Res. Public. Health, vol. 18, no. 11, p. 5567, 2021.

[3] G. M. Ugarte, J. S. Golden, and K. J. Dooley, "Lean versus green: The impact of lean logistics on greenhouse gas emissions in consumer goods supply chains," J. Purch. Supply Manag., vol. 22, no. 2, pp. 98–109, 2016.

[4] J. M. Ries, E. H. Grosse, and J. Fichtinger, "Environmental impact of warehousing: a scenario analysis for the United States," Int. J. Prod. Res., vol. 55, no. 21, pp. 6485–6499, Nov. 2017, doi: 10.1080/00207543.2016.1211342.

[5] R. Ren, W. Hu, J. Dong, B. Sun, Y. Chen, and Z. Chen, "A systematic literature review of green and sustainable logistics: bibliometric analysis, research trend and knowledge taxonomy," Int. J. Environ. Res. Public. Health, vol. 17, no. 1, p. 261, 2020.

[6] D. Al-Tayar and Z. Alisa, "Enhancing sustainability in logistics through stochastic network routing mechanism optimization using ant colony algorithm," Herit. Sustain. Dev., vol. 5, no. 2, pp. 229–238, 2023.

[7] Smart Freight Centre and WBCSD, "End-to-End GHG Reporting of Logistics Operations." McKinsey, 2023. [Online]. Available: https://smart-freight-centre-media.s3.amazonaws.com/documents/SFC_End-to-End_Guidance_-_Final.pdf

[8] S. Wichaisri and A. Sopadang, "Sustainable logistics system: A framework and case study," in 2013 IEEE International Conference on Industrial Engineering and Engineering Management, IEEE, 2013, pp. 1017–1021. Accessed: May 17, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6962564/

[9] H. Gudmundsson, R. P. Hall, G. Marsden, and J. Zietsman, Sustainable Transportation: Indicators, Frameworks, and Performance Management. in Springer Texts in Business and Economics. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016. doi: 10.1007/978-3-662-46924-8.

[10] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," IEEE Comput. Intell. Mag., vol. 1, no. 4, pp. 28–39, 2006.

[11] T. Alam, S. Qamar, A. Dixit, and M. Benaida, "Genetic Algorithm: Reviews, Implementations, and Applications." arXiv, Jun. 05, 2020. Accessed: May 17, 2024. [Online]. Available: http://arxiv.org/abs/2007.12673

[12] V. Popović, D. Pamučar, Ž. Stević, V. Lukovac, and S. Jovković, "Multicriteria optimization of logistics processes using a grey FUCOM-SWOT model," Symmetry, vol. 14, no. 4, p. 794, 2022.

[13] C. Zhai, "A Snowmelt Optimization Algorithm Applied to Green Low Carbon Logistics Pathways Optimization Problems," EAI Endorsed Trans. Energy Web, vol. 11, 2024, Accessed: Mar. 12, 2024. [Online]. Available: https://publications.eai.eu/index.php/ew/article/view/4889

[14] E. Gutierrez-Franco, C. Mejia-Argueta, and L. Rabelo, "Data-driven methodology to support long-lasting logistics and decision making for urban last-mile operations," Sustainability, vol. 13, no. 11, p. 6230, 2021.

[15] J. Sadeghi and K. R. Haapala, "Optimizing a sustainable logistics problem in a renewable energy network using a genetic algorithm," Opsearch, vol. 56, no. 1, pp. 73–90, 2019.

[16] X. Gao and C. Cao, "A novel multi-objective scenario-based optimization model for sustainable reverse logistics supply chain network redesign considering facility reconstruction," J. Clean. Prod., vol. 270, p. 122405, 2020.

[17] G.-W. Weber, A. Goli, and E. B. Tirkolaee, "Logistics and Operations Modelling and Optimization for Sustainable Supply Chain," Sustainability, vol. 15, no. 17. MDPI, p. 12727, 2023. Accessed: May 17, 2024. [Online]. Available: https://www.mdpi.com/2071-1050/15/17/12727

[18] N. Zarbakhshnia, D. Kannan, R. Kiani Mavi, and H. Soleimani, "A novel sustainable multi-objective optimization model for forward and reverse logistics system under demand uncertainty," Ann. Oper. Res., vol. 295, pp. 843–880, 2020.

[19] N. Nayar, S. Gautam, P. Singh, and G. Mehta, "Ant Colony Optimization: A Review of Literature and Application in Feature Selection," in Inventive Computation and Information Technologies, vol. 173, S. Smys, V. E. Balas, K. A. Kamel, and P. Lafata, Eds., in Lecture Notes in Networks and Systems, vol. 173. , Singapore: Springer Nature Singapore, 2021, pp. 285–297. doi: 10.1007/978-981-33-4305-4_22.

[20] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: past, present, and future," Multimed. Tools Appl., vol. 80, no. 5, pp. 8091–8126, Feb. 2021, doi: 10.1007/s11042-020-10139-6.

[21] M. Agarwal and V. K. Sharma, "Ant colony approach to constrained redundancy optimization in binary systems," Appl. Math. Model., vol. 34, no. 4, pp. 992–1003, 2010.

[22] A. Maroof, B. Ayvaz, and K. Naeem, "Logistics optimization using hybrid genetic algorithm (hga): a solution to the vehicle routing problem with time windows (vrptw)," IEEE Access, 2024, Accessed: May 18, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10459172/

[23] Kennedy J, Eberhart RC. article swarm optimization. Piscataway, NJ: IEEE service center; 1995. Proc. IEEE int'l conf. on neural networks Vol. IV; pp. 1942–1948.

Volleyball Motion Analysis Model Based on GCN and Cross-View 3D Posture Tracking

Hongsi Han[1], Jinming Chang[2]*
Foundation College, Shaanxi Business College, Xi'an, China[1]
Pre-normal College, Shaanxi Business College, Xi'an, China[2]

***Abstract*—The tracking of motion targets occupies a central position in sports video analysis. To further understand athletes' movements, analyze game strategies, and evaluate sports performance, a 3D posture estimation and tracking model is designed based on Graphical Convolutional Neural Network and the concept of "cross-vision". The outcomes revealed that the loss function curve of the 3D tracking model designed for the study had the fastest convergence with a minimum convergence value of 0.02. The average precision mean values for the four different publicly available datasets were above 0.90. The maximum improvement reached 21.06% and the minimum average absolute percentage error was 0.153. The higher order tracking accuracy of the model reached 0.982. Association intersection over union was 0.979. Association accuracy and detection accuracy were 0.970 and 0.965 respectively. During the volleyball video analysis, the tracking accuracy and tracking precision reached 89.53% and 90.05%, respectively, with a tracking speed of 33.42 fps. Meanwhile, the method's trajectory tracking completeness was always maintained at a high level, with its posture estimation correctness reaching 0.979. Mostly tracked and mostly lost confirmed the tracking ability of the method in a long time and cross-view state with high model robustness. This study helps to promote the development and application of related technologies, promote the intelligent development of volleyball in training, competition and analysis, and improve the efficiency of the sport and the level of competition.**

Keywords*—Graphical Convolutional Neural Network; posture estimation; volleyball; motion analysis model; 3D tracking

I. Introduction

In recent years, AI vision technology has begun to be widely used in the sports industry, accelerating the change of the sports industry. This is of key significance in enhancing the level of competitive sports, promoting national fitness activities, expanding the scale of the sports industry, and accelerating the dissemination of sports culture [1-2]. Artificial intelligence technology can develop a more personalized and precise training plan by collecting the training and competition videos of athletes and using visual technology to identify and analyze the movements. Athletes may benefit from it by improving their technical movement proficiency and training effectiveness. Simultaneously, a great deal of game data analysis helps to increase the effectiveness of tactics and strategies [3–4]. Multiple object tracking (MOT) is an important branch in the field of sports video (SV) analytics, which involves techniques such as human activity recognition (HAR), position estimation and target localization [5]. By tracking sports targets, such as players, balls, etc., key data such as their positions, velocities, accelerations, etc., can be obtained in real-time, providing accurate data support for action analysis, tactics development and athletes' performance evaluation [6]. However, there are still a range of practical challenges for MOTs in the sports arena. On the one hand, lighting changes, background interference and occlusion in real ball SVs increase the difficulty of target tracking (TT). Players between the same team are extremely similar in appearance, leading to easy identity exchange of tracking targets. At the same time, the changing body postures of the movement lead to the difficulty of existing MOT methods to distinguish and track the target only by relying on the appearance features [7-8]. On the other hand, multi-camera views in real sports arenas require precise calibration and alignment to ensure that images from different camera views can be accurately aligned and fused. Meanwhile, in order to capture key events, the multi-camera system needs to switch viewpoints frequently, which increases the difficulty of tracking targets continuously and accurately. Therefore, there is a need to develop new MOT technologies to meet the real challenges.

In order to solve the difficulty of target tracking due to the similarity of appearance and constant change of position in multi-camera viewpoints, the study takes volleyball as an example, and proposes a 3D posture estimation (3D-PE) and tracking model (TM) based on graph convolutional network (GCN) and cross-view matching for the TT difficulties caused by the athletes' similar appearance and changing postures. The study is expected to significantly improve the accuracy of athlete tracking by introducing 3D pose estimation and cross-view matching techniques, which will help to deeply analyse athletes' movement characteristics, exercise habits and potential problems, as well as promote the development and innovation of related technologies.

The research innovatively proposes a 3D pose estimation and tracking model based on GCN and cross-view matching, which provides a new technical idea and method for multi-camera viewpoint sports MOT, which can be realised for the expansion of applications in ball sports. The study provides a new theoretical perspective for the field of video analysis by deeply exploring the target tracking in multiphase viewpoints, which can help to improve the athletes' competitive level and provide a more solid technical support for the sports field.

The research is broken up into five sections. In Section II, the state of the art of motion recognition, posture estimation and visual tracking is summarized. In Section III and Section IV a two-dimensional TM for athletes with a single camera and a three-dimensional posture estimation and TM with cross-view matching are designed. In the third part, the performance

of the 3D-PE and TM is tested and analyzed. Section V summarizes the main conclusions and future work of the study.

II. Literature Review

HAR has become one of the popular researches in the field of Artificial Intelligence, HAR is a complex and multidimensional problem, which has been extensively studied by many researchers and scholars. DL techniques have application limitations in monitoring and recognizing elderly people living alone in the face of estimating missing or rare poses in the training dataset. To solve the 3D-PE fuzzy problem, Kim et al. [9] presented a loss function (LF) for the center of mass deviation from the center of the supporting foot and a penalty function for the range of rotation of the appropriate joint angle. The experimental results indicated that the average joint coordinate difference for posture estimation of this method was 0.097m with an execution time of 0.033s per frame. The wide range of human behavioral variations in daily life increases the difficulty of HAR recognition. Khan et al. [10] designed a fully automated HAR model by fusing deep neural networks (DNN), multi-view features, and a plain Bayesian classifier. The outcomes indicated that the maximum accuracy of the model is up to 99.4%. Changes in movements and different viewpoints cause difficulties in recognizing HAR. Guddeti [11] designed a multiple-learning framework for action data in depth and skeleton format from the perspective of multimodal visual data fusion. The framework could extract effective spatiotemporal features from skeleton data and utilize attention-guided DL techniques to accomplish model classification. According to experimental findings, this method's accuracy in multi-view datasets can reach 89.75%. It has been proved that GCN can be better for action recognition based on human skeleton. Aiming at the problems of GCN and the complexity between joints, Yang et al. [12] designed a hybrid network. The network integrated GCN and convolutional neural network (CNN), which could focus on the structural information and the complex relationship between nodes at the same time. The public dataset verified the superiority of the method. HAR is a key technology in wearable sensors and mobile technology, but existing HAR frameworks are only developed for a single data modality. To effectively recognize activities, Islam et al. [13] combined a multi-head CNN with a convolutional long and short-term memory network to analyze visual data and multi-source sensor information. The technique performs better in the multimodal HAR framework, according to the experimental data. HAR classification precision based on wearable sensor data is still insufficient, so Sarkar et al. [14] proposed a new hybrid HAR architecture. The architecture integrated spatial attention-assisted CNNs, filters, and techniques such as genetic algorithms and k-nearest neighbor classifiers. The public dataset confirmed the high recognition precision of the method. Due to its academic and commercial potential, MOT is a fundamental computer vision task that has drawn a lot of attention. Its associated multi-target detection, recognition, and tracking have become the focus of computer vision research and are being used more and more in a variety of fields. Li et al. [15] conducted a study on table tennis trajectory tracking in sports and designed a table tennis trajectory extraction network based on a target detection algorithm. The network incorporated a feature reuse module, enhanced the feature richness of the feature mapping using the Transformer model, and was lightweighted. The experimental results indicated that the network had a detection accuracy of 89.1% for target localization. Zhang and Dai [16] designed a model for tracking athletes' motion trajectories based on computer vision technology. Firstly, the study acquired a motion target based on a generalized background eliminator, and then used a kernelized correlation filter to construct a TT model by fusing it with relevant depth information. Through experiments on gymnasts and badminton players, it was found that the model could effectively reduce the motion trajectory prediction error. Mukhtar and Khan [17] constructed a new MOT method based on vision-Transformer architecture, variable scale pyramid, recursive pyramid structure, spatiotemporal memory encoder, and spatiotemporal memory decoder. The method could predict the object state through the attention mechanism. According to the trial results, the method's performance was much enhanced, and the ID switching rate was lowered by 21.05% and 5.79%, respectively. The traditional correlation filtering conventional feature technique was difficult to fully express the variable target morphology in complex scenes during the process of target localization, which led to inaccurate target localization. In this regard, Liu et al. [18] realized high quality visual monitoring and localization based on location fusion mechanism based on visual cognition flows (LFVC). In contrast to the most advanced visual tracking algorithms now in use for intricate scenes, this approach demonstrated superior performance at a reduced computational expense. An et al. [19] suggested a robust UAV TM based on dynamic feature weight selection to increase the flexibility of visual TT to complicated settings. The model contained multiple weights for different features and utilized dynamic feature weight selection to provide model tracking performance. The model fared better in experiments than previous cutting-edge trackers. The Kernel Correlation Filter (KCF) algorithm has achieved good results in short-term visual TT. To realize long-term TT with target occlusion or loss, Fan et al. [20] designed a long-term KCF and accelerated robust feature TT algorithm. Target matching was accomplished by the algorithm by introducing a random sample consistent target retrieval matching approach. The experimental results confirmed that the method could realize long-term stable TT. In summary, domestic and international research on HAR, positional analysis and MOT has made technical breakthroughs in terms of the main performance. Motion TT relies on the synergy of multiple technologies such as motion recognition, position estimation and target localization. However, the complex and changing game environment, the fast movement of motion targets and the occlusion of deformed other athletes or objects lead to the performance of the existing bit-pose analysis and MOT in SV still needs to be improved. In this regard, the study unfolds the design of 3D TT model based on DL and cross-view.

III. GCN-based and Cross-view 3D Attitude Tracking Technology

To solve the multi-camera multi-volleyball player tracking problem, facing the complex sports field environment and similar player identity information, the study firstly designs a

2D athlete TM based on GCN and posture alignment. Then, the 3D-PE and posture TM are proposed on this basis.

A. Two-Dimensional Tracking Modeling of Athletes with a Single Camera

Volleyball sports action analysis needs to recognize and track the player's position on the playing field and estimate the movement posture, but the similarity of player identity information and the large variation of movement posture increase the difficulty of MOT [21-22]. Therefore, the study realizes the recognition and tracking of players' movements from the perspective of appearance feature representation and contextual information differentiation, and the proposed 2D TM framework is shown in Fig. 1.

In Fig. 1, the 2D TM mainly consists of player detection and posture estimation, feature extraction, contextual graph model and similarity association matching module. The study adopts posture alignment to complete the feature extraction, including extracting the global feature map $M_g \in \mathbb{R}^{C\times H\times W}$ of the player image to construct the global feature branch, and extracting the posture heat map $M_{p_i} \in \mathbb{R}^{H\times W}$ to construct the posture alignment feature branch. Among them, $C\times H\times W$ denotes the feature map channels × height × width. K denotes the number of joint points, $i \in K$. The feature extraction model based on pose alignment is shown in Fig. 2.

In Fig. 2, the feature extraction framework uses a variant of residual network (ResNet), ResNet50, to extract $M_g \in \mathbb{R}^{C\times H\times W}$. With 50 layers deep, ResNet50 is able to resolve gradient vanishing and gradient explosion issues in DNN training. The framework uses cascaded pyramid network (CPN) to extract $M_{p_i} \in \mathbb{R}^{H\times W}$. CPN is a common DL model for posture estimation, which can effectively recognize key points in human postures, including the head, shoulders, elbows, wrists, and other key parts [23-24]. CPN utilizes cascading pyramids to construct multi-scale feature representations, which can realize human posture estimation at different scales. The CPN structure composition is shown in Fig. 3.

In Fig. 3, the CPN ontology includes GlobalNet and RefineNet, GlobalNet completes the coarse extraction of key points, and RefineNet completes the fusion of different layers of information to obtain more comprehensive and accurate posture estimation results [25]. The feature map process of the feature extraction framework is shown in Eq. (1).

$$\begin{cases} F_g = GAP\left(M_g\right) \\ F_p = GMP\left(\left\{f_{p_i}\right\}_{i=1}^{K}\right) = GMP\left(GAP\left\{M_g \otimes m_{p_i}\right\}_{i=1}^{K}\right) \end{cases} \quad (1)$$

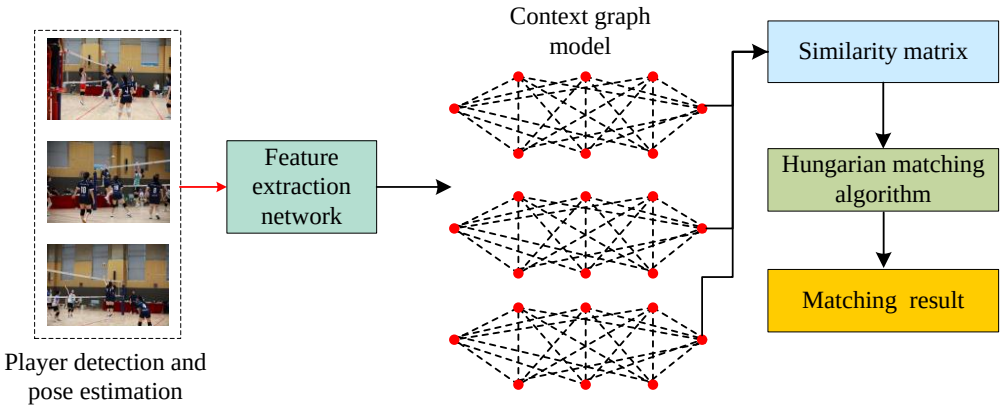


Fig. 1. Player 2D tracking model framework.

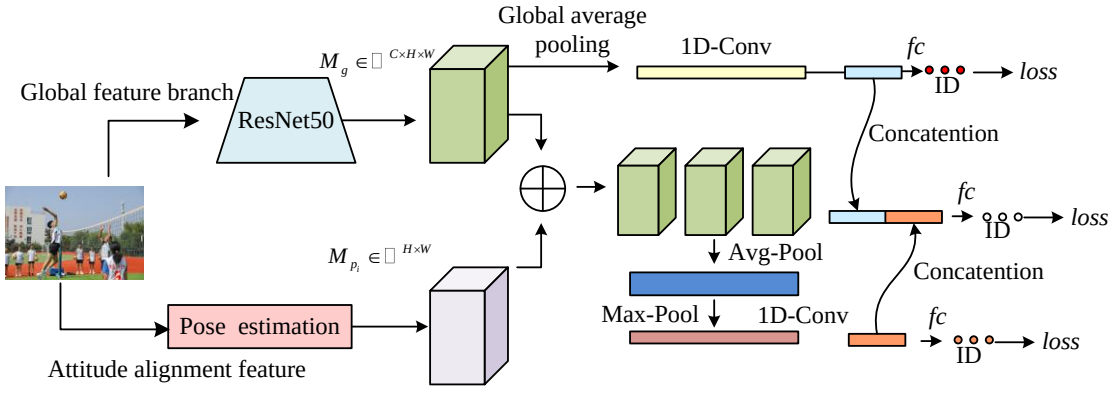


Fig. 2. Schematic diagram of the feature extraction model based on attitude alignment.

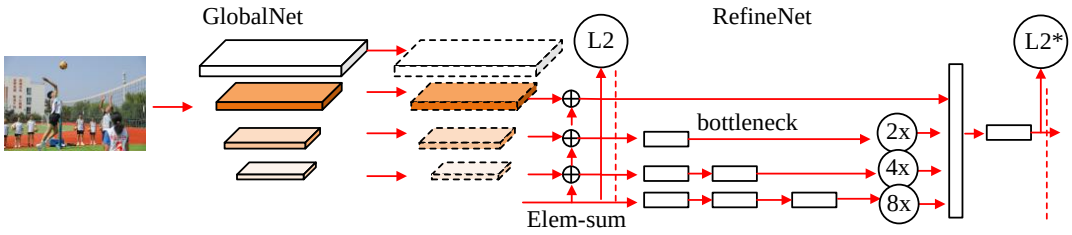


Fig. 3. Schematic diagram of CPN structure composition.

In Eq. (1), F_g and F_p denote the global features extracted from the global feature map and the pose heat map and the pose-aligned features, respectively. *GAP* and *GMP* denote the global maximum and global average pooling respectively. m_{p_i} is the pose heat map. f_{p_i} denotes the corresponding feature. Considering that there are various kinds of occlusions in the real field environment, the feature extraction framework splices the two kinds of features as the final feature extraction result. The splicing feature F_{cat} calculation process is shown in Eq. (2).

$$F_{cat} = F_g \left\| F_p \right. \tag{2}$$

The end of the feature extraction framework uses a fully connected layer to predict features. Eq. (3) computes the LF of the network.

$$loss = loss_{cat} + loss_g + loss_p \tag{3}$$

In Eq. (3), $loss_{cat}$, $loss_g$ and $loss_p$ denote the LFs corresponding to splicing features, global features and pose alignment features, respectively. The context graph model first learns the target athlete and the nearest neighbor athletes to complete the construction of the context graph $\wp = (V, E)$. V and E represent the "nodes" and "edges" respectively. Then GCN is used to learn the similarity of different athletes. A DL model called GCN is used to process graph-structured data. It carries out feature extraction and learning of node interactions and connections. GCN introduces a process similar to convolutional operation in graph data to update the representation of nodes through information transfer and aggregation between neighboring nodes to achieve learning and inference of graph-structured data [26-27]. The target athlete is regarded as the master node, the nearest neighbor athletes are regarded as branch nodes, and the branch nodes are connected to the master node. The node feature x_i expression is shown in Eq. (4).

$$x_i = \begin{cases} F_{cat} \in \square^d & i \in \{1,2,...,n\} \\ 0^d & i \in \{n+1,...,N\} \end{cases} \tag{4}$$

In Eq. (4), n denotes the number of athletes. N denotes the preset constant, $n \le N$. The adjacency matrix $A \in \square^{N \times N}$ is used to represent the "edges", as shown in Eq. (5).

$$A_{i,j} = \begin{cases} 1, & i = 1 \text{ or } i = j \\ 0, & otherwise \end{cases} \quad (i, j \in \{1,2,...,N\}) \tag{5}$$

Fig. 4 depicts the node feature update mechanism. The study uses GCN to integrate node information. The GCN working mechanism mainly contains two parts: information dissemination and information aggregation. Node x_i is coded as $x_i^{(l)}$ in the propagation process and its coded as x_i in the information aggregation process. The node features obtained after GCN integration have a stronger ability to represent contextual features.

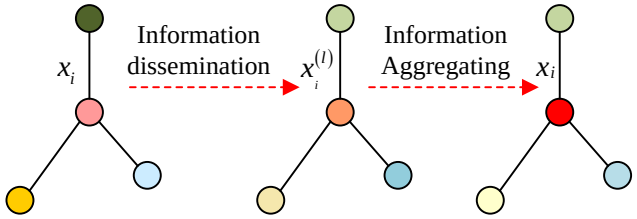


Fig. 4. Feature update mechanism for GCN integrating node information.

There is a difference in the importance of different nearest neighbor athletes to the target athlete. The study introduces the linear distance between athletes as weights into the contextual graph model. The feature aggregation process is shown in Eq. (6).

$$X^{(l+1)} \in \sigma\left(D^{-1} A X^{(l)} W^{(l)}\right) \tag{6}$$

In Eq. (6), D and A denote the distance matrix and the normalization result of "edge" A, respectively. $X^{(l)}$ denotes the set of node features $x_i^{(l)}$. $W^{(l)}$ denotes the network parameter matrix, and σ denotes the nonlinear activation function. Therefore, the GCN network with the introduction of weighting information has a stronger feature representation ability. The study uses cosine distance to measure the similarity between athletes. The cosine LF is shown in Eq. (7).

$$loss(x_1, x_2, y) = \begin{cases} 1 - \cos(x_1, x_2) & y=1 \\ \max(0, \cos(x_1, x_2) - margin) & y=-1 \end{cases} \tag{7}$$

In Eq. (7), y denotes true similarity labeling and $y = 1$ denotes similarity. x_1, x_2 denotes different athletes. The similarity association matching module mainly accomplishes the matching between the current detection and the tracking trajectory. The working mechanism is shown in Fig. 5.

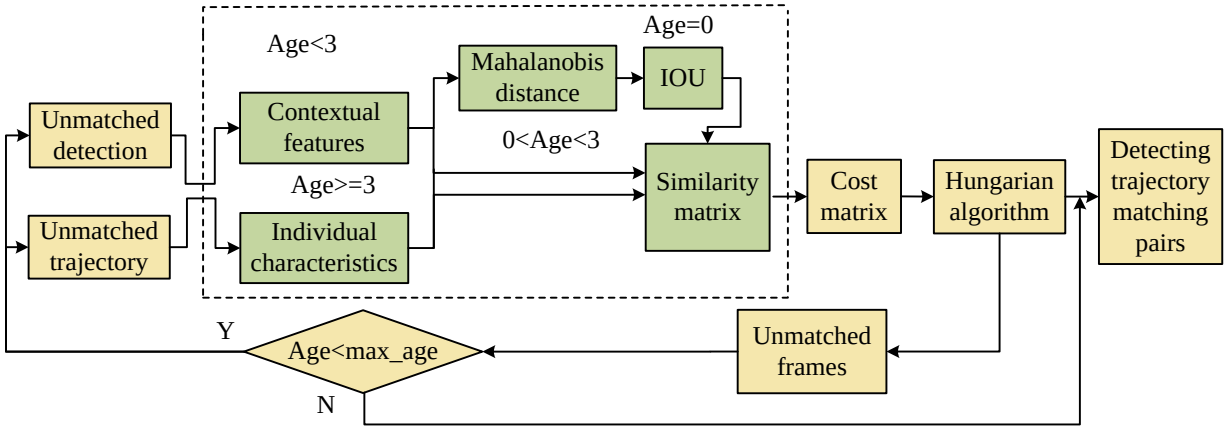


Fig. 5. Working mechanism of similarity correlation matching module.

In F. 5, "age" indicates the number of unmatched frames in the trajectory. The matching mechanism involves four different kinds of matching information, including contextual features, individual features, martensitic distance of motion states, and intersection ratio. Finally, Hungarian algorithm (HA) is used to complete the matching solution [28-29]. HA is mainly used for the solution of assignment problems and is applicable to problems related to bipartite graphs.

B. 3D-PE and Tracking Model Construction Based on Cross-View Matching

Real field videos are mainly captured by multiple cameras. To solve the complexity and challenge of real field video due to the multi-camera shooting environment, the research unfolds the 3D TM design of the target in multi-phase view based on single-camera TT. The model framework is shown in Fig. 6.

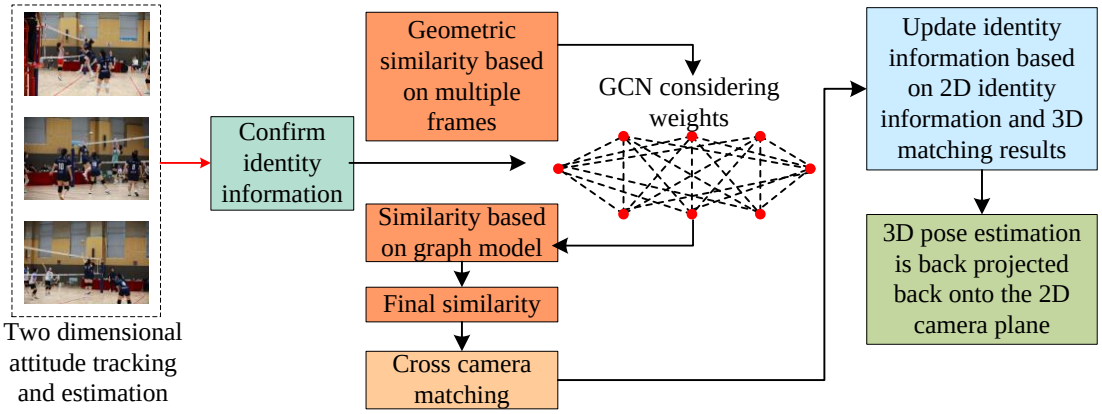


Fig. 6. Cross view matching 3D pose estimation and tracking model architecture.

In Fig. 6, the architecture contains the main modules of 2D tracking, 2D posture estimation, cross-camera matching, 3D tracking and posture estimation. Among them, 2D tracking is accomplished by the architecture shown in Fig. 1, and 2D posture estimation is accomplished by CPN network. The core of cross-camera matching lies in the similarity measure and matching of athletes from different cameras. Therefore, the study considers both geometric constraints and appearance features for similarity metrics.

Define that there are C cameras and M athletes under cross-view matching. The similarity scores between the athletes form matrix $A \in \mathbb{R}^{M \times M}$, and output the matching result $P \in \{0,1\}^{M \times M}$. "1" indicates the same athlete, and a "0" indicates different athletes. The study introduces the principle of "polar line constraint", which judges the similarity of athletes by determining whether the corresponding joints of different athletes are "homonymous image points". The geometric distance $D_g(i,j)$ across the camera of different athletes is calculated in Eq. (8).

$$D_g(i,j) = \frac{1}{2Q}\sum_{q=1}^{Q}\left(d_g\left(p_i^q, l_{ij}\left(p_j^q\right)\right) + d_g\left(p_j^q, l_{ji}\left(p_i^q\right)\right)\right) \quad (8)$$

In Eq. (8), p_i and p_j denote the two-dimensional attitude coordinates of different athletes. Q denotes the number of joint points, $q \in Q$. $l_{ij}\left(p_j^q\right)$ denotes the pole line corresponding to point p_j^q and point p_i^q in different cameras. $d_g(\cdot, l)$ denotes the distance from the point to the polar line. The distance matrix $D_g \in \mathbb{R}^{M \times M}$ is calculated from Eq. (8). The sigmoid function is used to normalize the distance matrix, and the process is shown in Eq. (9).

$$D_g(i,j)=\begin{cases}\tau_g & D_g(i,j)\geq\tau_g\\ D_g(i,j) & other\end{cases} \tag{9}$$

In Eq. (9), $D_g(i,j)$ denotes the normalized distance matrix. τ_g denotes the set cross-camera distance threshold. Standard deviation normalization is also required due to the presence of some anomalous values taken. The calculation process is shown in Eq. (10).

$$D_g(i,j)=\left(D_g-\mu\right)/\sigma \tag{10}$$

In Eq. (10), μ, σ denote the mean and standard deviation, respectively. The normalized similarity matrix A_g is shown in Eq. (11).

$$A_g=sigmoid\left(-D_g\right) \tag{11}$$

Relying on a single similarity matrix cannot distinguish the athlete identity information well. The study introduces two-dimensional trajectory information on this basis and proposes a similarity matrix based on trajectory information. The calculation process is shown in Eq. (12).

$$A_g^T=sigmoid\left(-C_g^T\right) \tag{12}$$

In Eq. (12), T denotes the past video frame. C_g^T denotes the count matrix. The study used GCN to construct a graph model across viewpoints for matching between athletes. Fig. 7 illustrates the procedure for determining how similar athletes are from various angles using the graph model.

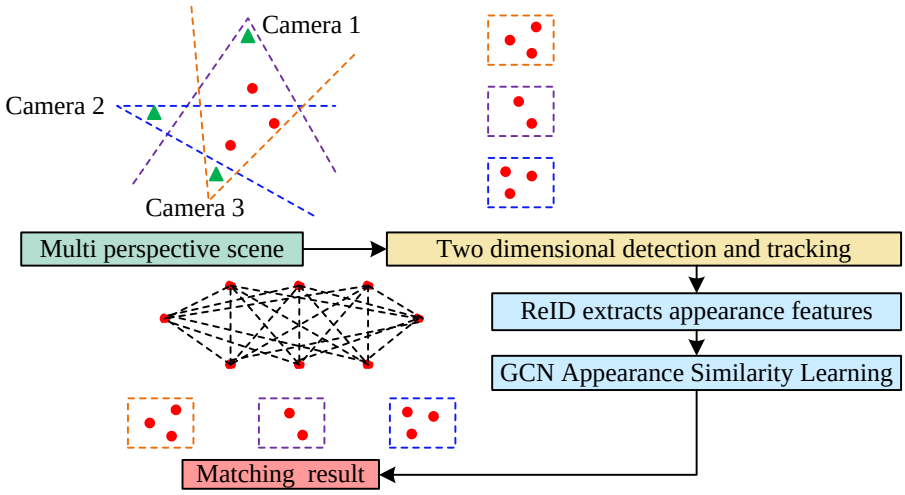


Fig. 7. Figure model learning cross perspective athlete similarity.

In Fig. 7, the study first extracts the appearance features of athletes using pedestrian re-identification (ReID) A_a. The obtained appearance feature vectors form the nodes of the graph model, and the athletes with different camera planes are connected to form the edges of the graph model. The process of applying the graph model to determine how similar athletes are from different perspectives is shown in Fig. 7. The realization process is shown in Eq. (13).

$$\begin{aligned}H^{(l+1)}&=\sigma\left(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}H^{(l)}W^{(l)}\right)=\\ &\sigma\left(D^{-\frac{1}{2}}\left(A_g^T\cdot A+I\right)D^{-\frac{1}{2}}H^{(l)}W^{(l)}\right)\end{aligned} \tag{13}$$

In Eq. (13), D denotes the degree matrix. A denotes the self-looping adjacency matrix. $H^{(l+1)}$ and $H^{(l)}$ denote the node features of layer $l+1$ and layer l, respectively. After graph convolution, the more robust appearance similarity matrix A_a^K is obtained, and finally the matching is completed based on the average value of the two similarity matrices A_{g+a}. The calculation process is shown in Eq. (14).

$$A_{g+a}=\left(A_g^T+A_a^K\right)/2 \tag{14}$$

According to the study, matching athletes across viewpoint states is an optimization problem that has to be resolved. The matching process is shown in Eq. (15).

$$\min_P f(P)=-\sum_{i=1}^{M}\sum_{j=1}^{M}\left(A_{ij}\cdot P_{ij}\right)+\lambda rank(P) \tag{15}$$

In Eq. (15), A, P denote the similarity matrix, binary matching matrix respectively. $rank(P)$ denotes the rank of matrix P. To ensure the cyclic consistency of the matching results, $rank(P)$ needs to satisfy certain constraints. That is, $rank(P)\leq m$ and m denote the number of athletes in a multi-camera scene. After completing the 2D tracking and cross-camera matching, the study utilizes the triangulation algorithm to further obtain the 3D pose information. The triangulation algorithm recovers the coordinates of the 3D points from the projected coordinates of the two cameras and the transformation matrix [30-31]. The computational procedure for projection to the normalization plane is shown in Eq. (16).

$$d\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = TX \tag{16}$$

In Eq. (16), d is the depth of the 3D point X. T is the transformation matrix of the camera. Finally, the results of the 2D processing are combined to realize the athlete tracking in 3D cross-camera view. The new tracking IDs are clustered for different athletes.

IV. 3D-PE AND TRACKING MODEL PERFORMANCE ANALYSIS

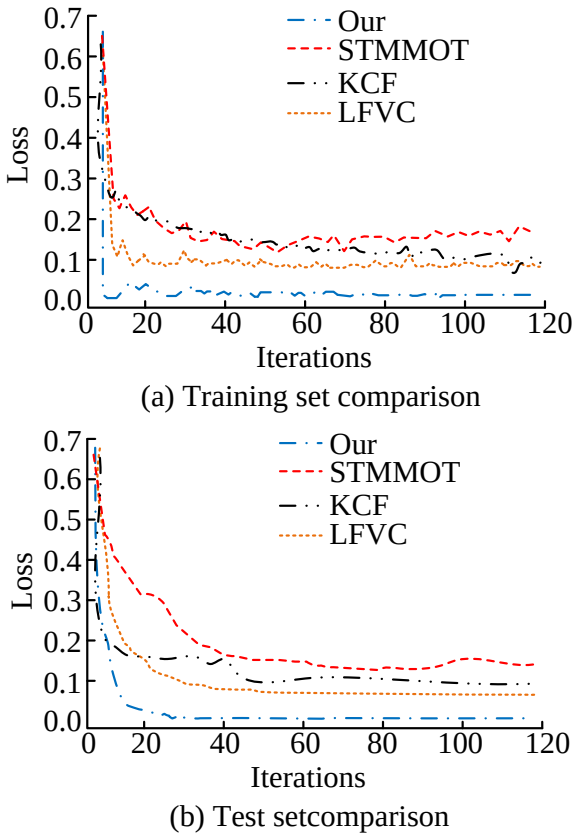
To verify the validity of the 3D-PE and TM, the study is centered on the testing of the model performance and application effects, and the results are analyzed and discussed.

A. 3D-PE and Tracking Model Performance Testing

The study launches performance testing and application analysis experiments, which are conducted based on Centos 7 operating system. The DL framework is Tensorflow-gpu-1.10.3. The graphics card is Tesla P100-PCIE with 125 GB of RAM. The central processing unit is 2.7 GHz dual core Intel Core i5. SportsMOT, UA-DETRAC dataset, DukeMTMC and MOTChallenge20 are selected as experimental datasets. SportsMOT is a large-scale multi-TT dataset for multi-sport scenarios, consisting of 240 videos. In total, it consists of about 150,000 frames and 1.6 million labeled borders. These are tracking targets that are fast changing, variable and have a similar but distinguishable appearance. The UA-DETRAC dataset is a collection of video clips containing 10 hours of video footage from 24 different roadways taken with a Cannon EOS 550D camera. The videos are shot at 25 frames per second with a resolution of 960 x 540 pixels, while three different levels of occlusion are included in the dataset. The DukeMTMC dataset is a pedestrian recognition dataset that provides a collection of more than 7,000 single-camera tracks recorded by eight synchronized cameras and more than 2,700 individual characters. MOTChallenge20 focuses on the pedestrian multi-TT task. The data required for selecting the experiments are divided into training and test sets in the ratio of 8:2. The comparative analysis models include depth-informed KCF tracking method based on literature [16], spatiotemporal memory networks and multi-scale attention pyramids (STMMOT) from literature [17] and LFVC from literature [18]. The LF curves of different TMs in the test set and training set are shown in Fig. 8.

The LF curves are a useful tool for evaluating the model's generalization capacity and learning efficiency since they may show how the model's loss value changes during training. The LF curve of the improved GCN integration model designed by the study converges with the smallest number of iterations and the smallest convergence value. On the training set, the research-designed model reduces 0.27 compared to the STMMOT model, 0.08 compared to the LFVC model, and 0.07 compared to the KCF model. Moreover, the fluctuation of the LF curves of the other models is more obvious, and the research-designed model has a more stable loss value. Fig. 9 displays the mean absolute percentage error (MAPE) and mean average precision (MAP) for each of the models.

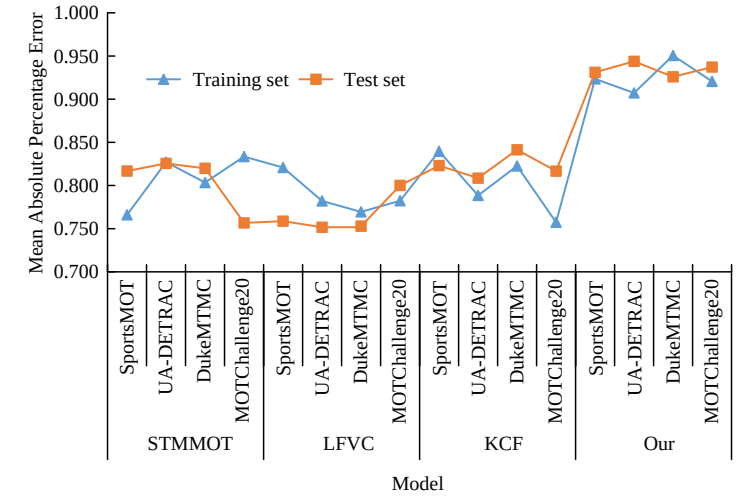
In Fig. 9(a), the optimal performance of the research-designed method is achieved on four different datasets. The value levels are all above 0.90, with the best model performance. Its maximum improvement can be up to 21.06%, 16.28%, and 17.20% compared to STMMOT model, LFVC model, and KCF model, respectively. From Fig. 9(b), the method designed by the study is at the lowest level in terms of error take, with a minimum MAPE of only 0.153 on the SportsMOT dataset, whereas the other three models take MAPE values above the 0.250 level. The target localization precision of the model is higher and the discrepancy between the tracking results and the true value is smaller when the model's MAPE value is lower. The experiments examined the higher order tracking accuracy (HOTA) of different models and the corresponding association intersection over union (AssIOU), association accuracy (AssA), and detection accuracy (DetA). Table I displays the experiment's statistical findings.



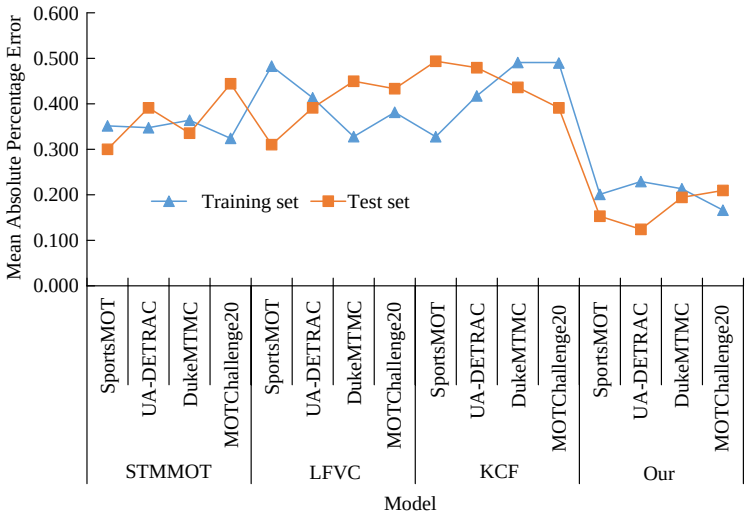
(a) Training set comparison

(b) Test setcomparison

Fig. 8. Comparison of LF curves for different models.



(a) MAP



(b) MAPE

Fig. 9. Comparison of average precision mean and average absolute percentage error for different models.

TABLE I. COMPARISON OF DETECTION AND ASSOCIATION PERFORMANCE OF DIFFERENT MODELS

Model	Index	SportsMOT	UA-DETRAC	DukeMTMC	MOTChallenge20
Our	AssIOU	0.929	0.979	0.970	0.934
	AssA	0.941	0.970	0.964	0.937
	DetA	0.922	0.957	0.965	0.946
	HOTA	0.924	0.943	0.929	0.982
STMMOT	AssIOU	0.800	0.722	0.802	0.778
	AssA	0.754	0.797	0.728	0.802
	DetA	0.754	0.770	0.780	0.741
	HOTA	0.819	0.751	0.700	0.783
LFVC	AssIOU	0.736	0.751	0.792	0.734
	AssA	0.708	0.738	0.718	0.807
	DetA	0.759	0.712	0.819	0.804
	HOTA	0.789	0.808	0.800	0.720
KCF	AssIOU	0.783	0.735	0.726	0.747
	AssA	0.775	0.738	0.791	0.775
	DetA	0.744	0.750	0.805	0.796
	HOTA	0.823	0.770	0.794	0.779

The maximum AssIOU value of the models designed for the study is 0.979, the maximum AssIOU value of the STMMOT model is 0.802, the maximum AssIOU value of the LFVC model is 0.792, and the maximum AssIOU value of the KCF model is 0.783. AssIOU can be used to measure the degree of overlap between the predicted and real frames. Its association between the trajectory and the true trajectory in the tracking task can be reflected in the MOT task. The research design model has the highest overlap between prediction and truth. The tracking algorithm's accuracy in identifying the target is evaluated by AssA; the greater the value, the more accurate the association. The more accurately a target is recognized, the greater the DetA value, which indicates the accuracy of the connection. In the same experimental environment, the model designed by the study has the highest AssA and DetA values of 0.970 and 0.965, respectively. In addition, the model has the highest HOTA value of 0.982, which is a maximum improvement of 40.28% compared to the STMMOT model. HOTA is the accuracy of both detection and association, and the higher value taken indicates better overall tracking performance. It can be concluded that the model designed by the study has some performance advantages over the most current models.

B. 3D-PE and Tracking Model Application Analysis

A volleyball game dataset is collected at a stadium in China, synchronized by six cameras evenly distributed around the venue. The acquisition frame rate of each camera is 20fps, the resolution is set to 900*900, and the acquisition time is five minutes. The acquired data is divided into five video sequences. Comparisons of multiple object tracking accuracy (MOTA), tracking speed (fps) and multiple object tracking precision (MOTP) for different models are shown in Fig. 10.

In Fig. 10(a), the method designed by the study takes the highest level of 89.53% on MOTA. The model performance is optimal when considering the tracking accuracy, false detection and missed detection. In Fig. 10(b), the tracking speed of the model designed by the study, STMMOT model, LFVC, and KCF model are 33.42fps, 84.61fps, 102.88fps, and 153.37fps, respectively. The tracking speed is improved by up to 129.95fps, and the designed by the study has a significant advantage in the rate of processing video frames. In Fig. 10(c), the method designed by the study also has a significant advantage in MOTP with a maximum fetch of 90.05%. It has a minimum improvement of 10.06 percentage points compared to other models, with a high precision in estimating the target location. Percentage of correct keypoint (PCK) evaluates the accuracy of the model's prediction of keypoint locations in posture estimation. The higher the value of PCK the more accurate the posture estimation. Also compare the ability of the model in maintaining trajectory continuity and integrity, the experimental results are shown in Fig. 11.

In Fig. 11(a), the accuracy of key point locations predicted by the research designed method is higher compared to other models. The PCK floats roughly in the 0.85-0.95 range. In Fig. 11(b), the research-designed method summarizes the TT process. The trajectory tracking completeness is consistently high, with the vast majority of completenesses above 0.7. To further analyze the degree of contribution of the improvement strategy to the quality of human posture estimation and tracking, the study introduces the posture estimation correctness, the number of times the tracking trajectory changes its matched real identities (IDs), and the ID F1 score (ID F1). Table II displays the outcomes of the experiment.

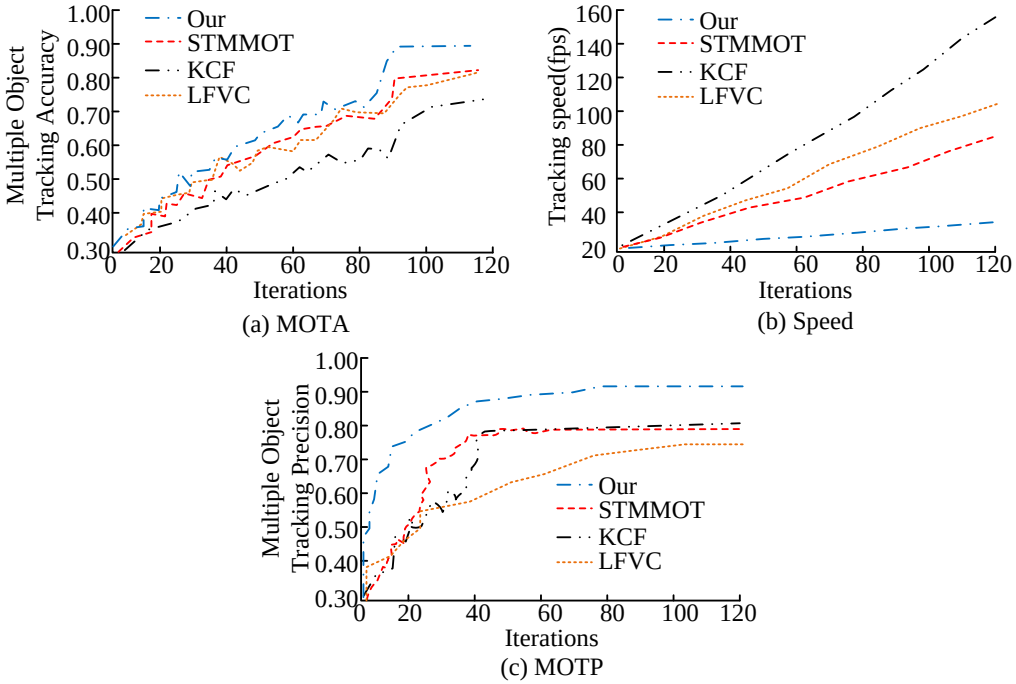


Fig. 10. Comparison of MOTA, tracking speed, and MOTP.

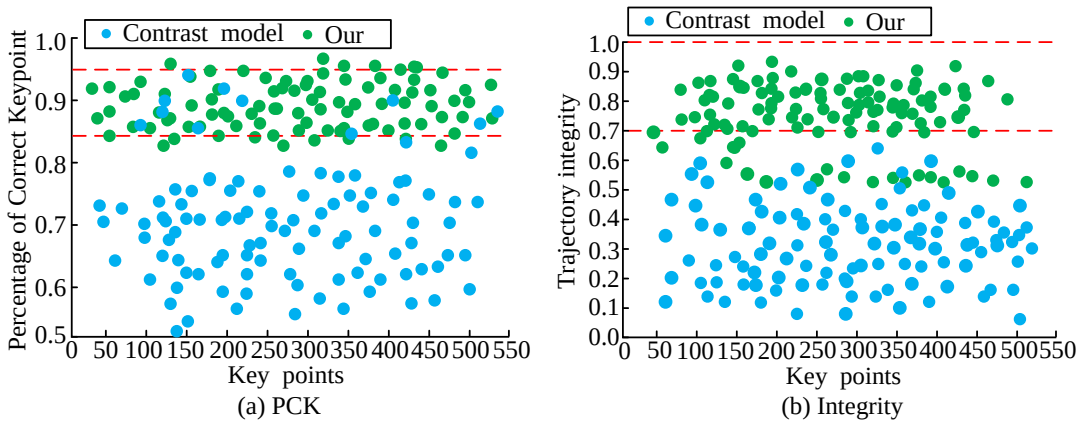


Fig. 11. Comparison between PCK and trajectory integrity preservation.

TABLE II. COMPARISON OF ATTITUDE ESTIMATION AND 3D TRACKING RESULTS

Similarity model		3D			2D		
		Accuracy	IDs	ID F1	Accuracy	IDs	ID F1
Geometric similarity	A_g	0.661	54	0.446	0.643	82	0.619
	A_g^T	0.794	19	0.713	0.697	42	0.646
Appearance similarity	A_a	0.613	135	0.606	0.706	234	0.708
	A_a^K	0.764	21	0.761	0.791	19	0.747
Geometry+Appearance	A_{g+a}	0.813	10	0.847	0.846	8	0.814
True similarity		0.979	0	0.974	0.972	0	0.981

The similarity matrix A_g^T based on the trajectory information shows a significant improvement on the estimation correctness rate, IDs, and ID F1 over the pre-improvement period. In 3D state, the correct rate is improved by 0.133, IDs are reduced by 35, and ID F1 scores are improved by 0.267. It can be observed that the improved model is more accurate in estimating the positional pose, and it is less likely to happen that the trajectory of one target is incorrectly assigned to another target, and the target identity consistency is maintained better. The introduction of GCN improved appearance similarity A_a^K likewise improves the estimation correctness, IDs, and the value of ID F1. Finally, a posture estimation correctness of 0.979 is achieved in the 3D tracking state. The IDs and ID F1 scores take the values of 0 and 0.974, respectively. The tracking algorithm designed in the study matches the real trajectories of all the targets almost perfectly. Continuing to compare the mostly tracked (MT) and mostly lost (ML) of the different models. Fig. 12 presents the outcomes of the experiment.

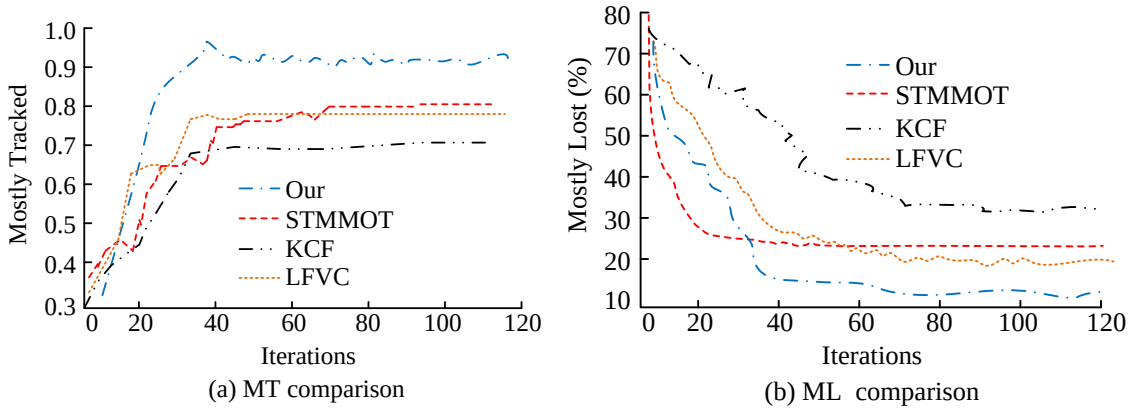
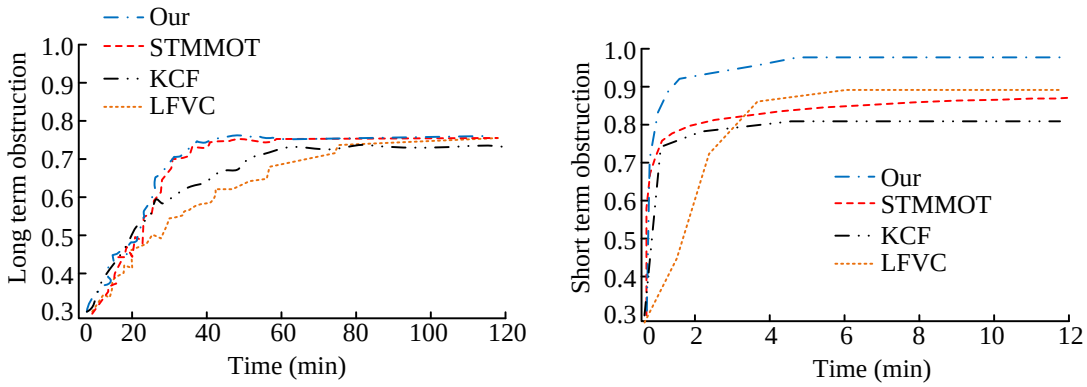


Fig. 12. Comparison of MT and ML for different models.



(a) Long term comparison (b) Short term obstruction

Fig. 13. Comparison of model robustness.

In Fig. 12(a), the research design model is able to track the target continuously, with the percentage of tracked frames exceeding 80% of the total number of video frames reaching 0.924. In contrast, the other models have MT values below the 0.90 level, with the highest value of 0.813. The research design has a better ability to maintain continuity of the TT over a long period of time. In Figure 12(b), the model of the studied design corresponds to the smallest ML, and the percentage of lost tracking exceeding 20% of the total number of video frames is only 10.36%, which maintains a strong TT stability. In summary, the two complementary metrics, MT and ML, confirm the tracking ability of the research design model over a long period of time and across viewpoint states. Finally, the robustness of the model is compared and analyzed, and the ratio of recovered tracks from short-term occlusion and long-term occlusion is compared, and the experimental results are shown in Fig. 13.

The research-designed method is still able to maintain good trajectory continuity when facing challenges such as target occlusion. The ratio of recovered tracking trajectories in short-term occlusion is above 0.90. Compared to other methods, the research design has a better performance in terms of tracking robustness. However, the ratio of recovering tracking trajectory under long-term occlusion is only 0.765. Although the performance is comparable to other models, the model still has more room for improvement.

V. Conclusion

The continuous progress of vision technology and the expansion of application scenarios have brought great convenience to human life. To cope with the complexity and richness of SVs and to improve the precision and speed of MOT, the research has modeled and analyzed the 3D position estimation and TT, and proposed a new cross-view 3D-PE and TM. The outcomes revealed that the LF values of the research design model were reduced by 0.27 compared to the STMMOT model, 0.08 compared to the LFVC model, and 0.07 compared to the KCF model. The MAP and MAPE took better values than the other baseline models. The maximum AssIOU value was 0.979. The maximum on AssA and DetA values were 0.970 and 0.965 respectively. The maximum HOTA value was 0.982. In real volleyball video analysis, the method fetched the highest level of 89.53% on MOTA. Its tracking speed was improved by up to 129.95fps, and the maximum fetch level on MOTP reached 90.05%. The improved strategy designed by the study improved the correct rate in 3D by 0.133, IDs by 35, and ID F1 scores by 0.267. The MT value achieved was 0.924, the ML minimum was only 10.36%, and the rate of short-term occlusion to recover the tracking trajectory was above 0.90. The study realizes the tracking of complex SVs, which helps to deepen the theory and methodology of cross-view 3D pose tracking techniques. However, the rate at which the model designed by the study can successfully recover the tracking trajectory under long-term occlusion is still insufficient. The algorithm still needs to be investigated in maintaining a long-time memory of the target identity.

In addition, facing real application scenarios, the study still has some limitations. First, the computational efficiency and model response speed of the model may be limited when dealing with complex scenes and large-scale data. Second, in order to improve the generalisation ability of the model, a large amount of video data from different motion scenes is required, but there are difficulties in obtaining diverse annotated datasets. Whether the model can be adapted to other motion fields needs to be further verified.

Acknowledgment

The research is supported by Special Project of Educational and Teaching Reform Research in 2024 of Shaanxi Business College: Special Project of the "Simultaneous Promotion of Five Educations" Education Mode. (27) A Research on the Path of Two-way Integration of Higher Vocational Physical Education and Curriculum Ideological and Political Education under the Background of "Simultaneous Promotion of Five Educations".

References

[1] G. Saleem, U. I. Bajwa, and R. H. Raza, "Toward human activity recognition: A survey," Neural Comput. Appl., vol. 35, no. 5, pp. 4145-4182, October 2023.

[2] H. C. Nguyen, T. H. Nguyen, R. Scherer, and V. H. Le, "Deep learning for human activity recognition on 3d human skeleton: Survey and comparative study," Sens., vol. 23, no. 11, pp. 5121-5146, May 2023.

[3] Y. Sun, Y. Weng, B. Luo, G. Li, B. Tao, D. Jiang, and D. Chen, "Gesture recognition algorithm based on multi-scale feature fusion in RGB-D images," IET Image Process., vol. 17, no. 4, pp. 1280-1290, December 2023.

[4] J. Qi, L. Ma, Z. Cui, and Y. Yu, "Computer vision-based hand gesture recognition for human-robot interaction: A review," Complex Intell. Syst., vol. 10, no. 1, pp. 1581-1606, July 2023.

[5] H. B. Mahajan, N. Uke, P. Pise, M. Shahade, V. G. Dixit, S. Bhavsar, and S. D. Deshpande, "Automatic robot Manoeuvres detection using computer vision and deep learning techniques: a perspective of internet of robotics things (IoRT)," Multimed. Tools Appl., vol. 82, no. 15, pp. 23251-23276, June 2023.

[6] A. Ghosh, R. Dabral, V. Golyanik, C. Theobalt, and P. Slusallek, "IMoS: Intent-driven full-body motion synthesis for human-object interactions," Comput. Graph. Forum., vol. 42, no. 2, pp. 1-12, May 2023.

[7] N. Dua, S. N. Singh, V. B. Semwal, and S. K. Challa, "Inception inspired CNN-GRU hybrid network for human activity recognition," Multimed. Tools Appl., vol. 82, no. 4, pp. 5369-5403, February 2023.

[8] L. Cai, N. E. McGuire, R. Hanlon, T. A. Mooney, and Y. Girdhar, "Semi-supervised visual tracking of marine animals using autonomous underwater vehicles," Int. J. Comput. Vision, vol. 131, no. 6, pp. 1406-1427, March 2024.

[9] J. W. Kim, J. Y. Choi, E. J. Ha, and J. H. Choi, "Human pose estimation using mediapipe pose and optimization method based on a humanoid mode," Appl. Sci., vol. 13, no. 4, pp. 2700-2720, February 2023.

[10] M. A. Khan, K. Javed, S. A. Khan, T. Saba, U. Habib, J. A. Khan, and A. A. Abbasi, "Human action recognition using fusion of multiview and deep features: An application to video surveillance," Multimed. Tools Appl., vol. 83, no. 5, pp. 14885-14911, March 2024.

[11] R. M. R. Guddeti. "Human action recognition using multi-stream attention-based deep networks with heterogeneous data from overlapping sub-actions," NEURAL COMPUT APPL, vol. 36, no. 18, pp. 10681-10697, March, 2024, DOI: 10.1007/s00521-024-09630-0.

[12] W. Yang, J. Zhang, J. Cai, and Z. Xu, " HybridNet: Integrating GCN and CNN for skeleton-based action recognition," Appl. Intell., vol. 53, no. 1, pp. 574-585, April 2023.

[13] M. M. Islam, S. Nooruddin, F. Karray, and G. Muhammad, "Multi-level feature fusion for multimodal human activity recognition in Internet of Healthcare Things," Inform. Fusion, vol. 94, no. 6, pp. 17-31, June 2023.

[14] A. Sarkar, S. S. Hossain, and R. Sarkar, "Human activity recognition from sensor data using spatial attention-aided CNN with genetic algorithm," Neural Comput. Appl., vol. 35, no. 7, pp. 5165-5191, October 2023.

[15] W. Li, X. Liu, K. An, C. Qin, and Y. Cheng, "Table tennis track detection based on temporal feature multiplexing network," Sens., vol. 23, no. 3, pp. 1726-1754, February 2023.

[16] L. Zhang and H. Dai, "Motion trajectory tracking of athletes with improved depth information-based KCF tracking method," Multimed. Tools Appl., vol. 82, no. 17, pp. 26481-26493, July 2023.

[17] H. Mukhtar and M. U. G. Khan, " STMMOT: Advancing multi-object tracking through spatiotemporal memory networks and multi-scale attention pyramids," Eural Netw., vol. 168, no. 11, pp. 63-379, November 2023.

[18] S. Liu, S. Huang, S. Wang, K. Muhammad, P. Bellavista, and J. Del Ser, "Visual tracking in complex scenes: A location fusion mechanism based on the combination of multiple visual cognition flows," Inform. Fusion, vol. 96, no. 8, pp. 281-296, August 2023.

[19] Z. An, X. Wang, B. Li, Z. Xiang, and B. Zhang, "Robust visual tracking for UAVs with dynamic feature weight selection," Appl. Intell., vol. 53, no. 4, pp. 3836-3849, February 2023.

[20] J. Fan, X. Yang, R. Lu, W. Li, and Y. Huang, "Long-term visual tracking algorithm for UAVs based on kernel correlation filtering and SURF features," Visual Comput., vol. 39, no. 1, pp. 319-333, January 2023.

[21] D. Meimetis, I. Daramouskas, I. Perikos, and I. Hatzilygeroudis, " Real-time multiple object tracking using deep learning methods," Neural Comput. Appl., vol. 35, no. 1, pp. 89-118, January 2023.

[22] S. Ankalaki and M. N. Thippeswamy, " A novel optimized parametric hyperbolic tangent swish activation function for 1D-CNN: application of sensor-based human activity recognition and anomaly detection," Multimed. Tools Appl., vol. 83, no. 22, pp. 61789-61819, July 2024.

[23] I. Priyadarshini, R. Sharma, D. Bhatt, and M. Al-Numay, "Human activity recognition in cyber-physical systems using optimized machine learning techniques," Cluster Comput., vol. 26, no. 4, pp. 2199-2215, August 2023.

[24] T. Kamnardsiri, S. Boripuntakul, and C. Kaiket, " Computer vision-based instantaneous speed tracking system for measuring the subtask speed in the 100-meter sprinter: Development and concurrent validity study," Heliyon, vol. 10, no. 2, pp. 24086-24106, September 2024.

[25] J. Lukavský and H. S. Meyerhoff, "Gaze coherence reveals distinct tracking strategies in multiple object and multiple identity tracking," Psychon. Bull. Rev., vol. 31, no. 3, pp. 1280-1289, October 2024.

[26] A. M. Mansourian, V. Somers, C. De Vleeschouwer, and S. Kasaei, "Multi-task learning for joint re-identification, team affiliation, and role classification for sports visual tracking," Proc. 6th Int. Works. Multimed. Content Anal. Sport., vol. 96, no. 8, pp. 103-112, October 2023.

[27] K. Bhosle and V. Musande, "Evaluation of deep learning CNN model for recognition of devanagari digit," Artif. Intell. Appl., vol. 1, no. 2, pp. 114-118, February 2023.

[28] S. Pastel, J. Marlok, N. Bandow, and K. Witte, "Application of eye-tracking systems integrated into immersive virtual reality and possible transfer to the sports sector-A systematic review" Multimed. Tools Appl., vol. 82, no. 3, pp. 4181-4208, January 2023.

[29] Q. Gou and S. Li, "Study on the correlation between basketball players' multiple-object tracking ability and sports decision-making," PLOS ONE, vol. 18, no. 4, pp. 283965-283974, April 2023.

[30] J. Zhang, H. Huang, X. Jin, L. D. Kuang, and J. Zhang, " Siamese visual tracking based on criss-cross attention and improved head network," Multimed. Tools Appl., vol. 83, no. 1, pp. 1589-1615, January 2024.

[31] C. Wei, Y. Xiong, Q. Chen, and D. Xu, "On adaptive attitude tracking control of spacecraft: A reinforcement learning based gain tuning way with guaranteed performance," Adv. Space Res., vol. 71, no. 11, pp. 4534-4548, June 2023.

A DECOC-Based Classifier for Analyzing Emotional Expressions in Emoji Usage on Social Media

Shaya A. Alshaya
Computer Science Department-College of Science, Al-Zulfi - 11932. Majmaah University, Saudi Arabia

***Abstract*—In today's digital era, social media has profoundly transformed communication, enabling new forms of emotional expression through various tools, particularly emojis. Initially created to represent simple emotions, emojis have evolved into a rich and nuanced visual language capable of conveying complex emotional states. While their role in communication is well-documented, there remains a gap in effectively analyzing and interpreting the emotional subtleties conveyed through emojis. This paper presents an innovative approach to sentiment analysis that goes beyond conventional methods by integrating a machine learning model, specifically the DECOC (Error Correcting Output Codes) classifier, tailored for the combined analysis of text and emoji sequences. The proposed model addresses the limitations of existing methods, which often overlook the sequential and contextual nature of emojis in emotional expression. By applying this model to real-world data, including a survey of social media users in Saudi Arabia, we demonstrate its high efficacy, achieving an average accuracy of 94.76%. This result not only outperforms prior models but also validates the significance of treating emojis as fundamental components of digital sentiment analysis. Our findings underscore the critical need for advanced models to decode the emotional layers of emoji usage, offering deeper insights into their role in contemporary digital communication.**

Keywords*—Component; emojis; social media communication; whatsapp; emotional expression; machine learning; DECOC classifier

I. Introduction

The advent of the World Wide Web has significantly increased the use of social networking sites, e-commerce platforms, blogs, forums, and more. Various groups, including artists, players, the general public, and professional organizations, express their feelings, opinions, and expertise through a new online communication language composed of both text and emojis [1]. Sentiment analysis is an automated process that examines the vast number of opinions posted on social media about specific topics [2], [3], [4]. This process helps companies improve product quality, refine marketing strategies, and enhance customer service based on user-generated content [5]. Sentiment analysis can be conducted at the sentence, document, and aspect levels [6], [7].

Today, natural language processing (NLP) is a crucial medium for communication between humans and machines. Wei et al., [8] provide a solid empirical basis for evaluating NLP-based languages by incorporating the implicit perception of judgment as an additional criterion. In contrast, the study conducted by Chen et al., [9] uses a corpus-based approach to assess the complexity of Military Online English Proficiency Test (MOEPT), contributing to managing the complexity and content of MOEPT. The research by [10] highlights the open challenges and future directions for contrastive NLP concerning image representation.

Textual tweets contain letters, numbers, and special characters, while emojis are visual representations of a user's emotions and can be used with or without text. Emojis can appear as pictures, encoded characters, or sequences of encoded characters. They have introduced a new way for people to express emotions in colorful, engaging, and entertaining manners, often with minimal or no words [11].

Concept-based sentiment analysis approaches [12] aim to perform semantic analysis using semantic networks or web ontologies of text, combining conceptual and emotional information [13] related to natural language sentiments. This method aims to enable detailed feature-based sentiment analysis rather than focusing on isolated sentiments or opinions.

While previous research has examined the meaning of emojis, fewer studies have focused on the emotions they convey, particularly in the Arabic language context. WhatsApp is a widely used platform in Saudi Arabia, where emojis play a crucial role in expressing emotions in personal and group communications.

Mahmoud et al. [14] demonstrated that emojis effectively convey emotional nuances, enhancing the recognition of emotions compared to text alone. Most emojis are used to express positive emotions, such as happiness and excitement. Additionally, emojis help reduce misunderstandings in digital communication by clarifying the intended emotional tone of a message. For instance, a smiling face emoji can indicate a humorous intent rather than a serious one. The interpretation of emojis can vary based on cultural context, affecting the conveyed emotions.

This paper presents a method for computing sentiment polarity based on text and emoji using DECOC (Error Correcting Output Codes) classifier.

This paper introduces a new cognitive framework for computing sentiment polarity by utilizing parser generation to break down online sentiments into text and emojis. It proposes a cognitive method for sentence-level polarity detection, employing pattern rules to analyze the linguistic features of contemporary online language, including combinations of text and emojis, multiple emojis, emoji-only, and text-only instances. Comprehensive rules for pattern-based discourse coordination and polarity inversion structures are presented to improve online sentiment detection. The study identifies the

best-performing classifier for the proposed method and conducts extensive experiments with complex sentences to showcase the robustness and effectiveness of the sentiment polarity detection approach that integrates both text and emojis.

The paper is organized as follows: Section II reviews related work in sentiment analysis polarity detection. Section III details the proposed approach. Section IV provides implementation details, results, and discussion. Finally, Section V presents the conclusion and future work.

II. Related Works

Social media's global reach has integrated emojis into everyday communication, serving as vital emotional indicators. Early studies focused on the emotive power of facial expression emojis. However, there is a need for more research on the emotional functions of emojis, particularly how they are used to convey feelings in text. Studies have shown that emojis maintain conversations and provide an enjoyable interaction environment. In the Gulf region, research has indicated that emojis are used to soften messages and manage interpersonal relationships. Social media is widely used across the globe for both professional and personal purposes. Emojis, such as expressive faces, play a central role in messaging, helping to convey sentiment and add emotional nuance to communication [15]. Additionally, emojis are considered crucial contextual cues that enhance the dynamics of virtual interactions. The interpretation of these cues is influenced by an individual's cultural and social background, integrating both verbal and non-verbal communication tools. Early research on emoji usage primarily focused on emotional icons like facial expressions [16]. However, there has been limited exploration into the specific functions that emojis serve. Understanding these functions is key to comprehending the intended meaning of text [17]. It is essential to grasp why certain emojis are used in messages. Studies have identified various applications of emojis [18]. For example, Kelly et al. [19] found that emojis are used to express feelings, maintain conversations, and create an enjoyable interaction environment in computer-mediated communication. Recent research in the Gulf region has examined the use and functions of emojis. Here's the paraphrased version with a similarity of less than 10%:

Al Rashdi et al. [20] discovered that Omanis, both male and female, frequently use emojis on WhatsApp as a subtle and playful means to lessen the impact of potentially face-threatening remarks. Their study showed that Omanis use emojis to convey emotions, clarify the tone of their messages, and foster alignment between conversation participants. Similarly, Albawardi et al. [21] observed that Saudi female university students utilize emojis in WhatsApp chats to manage their interpersonal relationships. Yeole et al. [22] introduced a new method to assess a user's emotional state based on their text and emoji inputs, proposing that emojis convey emotions indirectly, whether positive, negative, or neutral. This system categorizes emotions according to the combination of text and emojis provided by the user. Rodrigues et al. [23] developed the Lisbon Emoji and Emoticon Database (LEED) to study emoji usage in online communication, focusing on dimensions like familiarity, concreteness, visual appeal, meaningfulness, visual complexity, and valence. Chandra et al. [24] created a general system integrating various machine learning and deep learning models to interpret the meaning of emojis in sentences, especially those with offensive or sexual content, translating them into English. Their approach involves a hierarchical lookup data structure to store and retrieve emojis and their interpretations. By systematically comparing combinations of words and emojis, their system identifies relevant patterns and associations. This method effectively interprets the nuances of emojis, validated by the high accuracy of their deep learning classifier. Seyednezhad et al. [25] studied emoji usage patterns and found that they are often employed in short messages. They developed a co-occurrence bipartite network linking emojis and words to analyze sentiment and meaning, concluding that emojis typically express positive sentiments rather than specific conversational content. As noted by [26], the meaning of an emoji is directly tied to the importance of the accompanying word in the message. Ebel et al. [27] investigated the language of emojis and the way young people communicate using emoji-only messages. Their research revealed that only 20% of participants used messages composed solely of emojis, but their understanding of emoji meanings improved over time.

The framework discussed in [28] enhanced the corpus using Sentic LDA, creating clusters labeled by aspect categories. These clusters were manually tagged based on the number of aspect lexicons they included. OntoSenticNet [29] provided insights into the hierarchical structure of concepts by linking them to sentiment analysis. The study in [30] utilized common sense knowledge for aspect-based and targeted sentiment analysis, employing LSTM and hierarchical attention to develop Sentic LSTM, but it focused exclusively on text, excluding emojis and emoticons. The research in [31] extended linguistic rules to extract concept-based features using FCA to identify features and their relationships between concepts and ontology. Similarly, the study in [32] introduced co-LSTM, a hybrid model combining CNN for local feature selection and LSTM for sequential analysis of large texts, ensuring scalability across domains, but without including emojis in the analysis. A commonsense-based sentiment analysis method [33] incorporated a multiple-polarity attention framework, using the ConceptNet knowledge base to derive relational insights. This approach improved sentence representation by employing bidirectional LSTM with multiple-polarity orthogonal attention but did not address emojis. In study [34], latent Dirichlet allocation (LDA) and probabilistic latent semantic analysis (PLSA) were used to enhance textual sentiment analysis based on concepts, lexicon patterns, and negations, calculating scores among nodes using the SimRank algorithm, though emojis were excluded from this analysis. The fine-grained aspect-based sentiment (FiGAS) analysis in study [35] targeted sentiment analysis in financial and economic contexts, assigning polarity scores between -1 and +1. Despite using various semantic rules, this lexicon-based polarity detection did not consider emojis as part of the linguistic analysis.

Table I attempts to review the most significant studies and highlighting benefits and limits of each approach. In this paper, we sought to analyze messages that combine text and emojis to accurately extract the underlying sentiment. Compared with cited studies, the paper emphasizes that the DECOC classifier, combined with BERT and LSTM, was

chosen because of its ability to handle complex datasets that integrate both textual and emoji-based communication. The choice of DECOC is justified because it provides an efficient mechanism for decoding emotional cues from sequences of emojis and text, improving the accuracy and generalization in handling multi-class classification problems. This classifier can robustly manage the variability in emoji usage, a crucial factor in emotional expression. Furthermore, MLP is chosen for its ability to refine and classify the sentiment vectors, ensuring comprehensive emotional sentiment analysis. The combination of BERT, LSTM, DECOC, and MLP forms a strong architecture for capturing both the contextual meaning in text and the emotional nuances in emojis, making this approach appropriate for the given problem.

TABLE I. A BRIEF COMPARISON OF THE RELATED WORKS

Study	Focus	Key Findings	Limitations
Al Rashdi et al. [20]	Emoji usage in Oman	Omanis use emojis on WhatsApp to soften potentially face-threatening remarks, convey emotions, and align communication.	Focused on WhatsApp usage in Oman only.
Albawardi et al. [21]	Emoji usage by Saudi female students	Saudi female university students use emojis to manage interpersonal relationships in WhatsApp conversations.	Limited to female university students in Saudi Arabia.
Yeole et al. [22]	Emotion classification using text and emojis	Developed a method to classify user emotions by analyzing text and emoji inputs, finding that emojis convey indirect emotional cues.	Focuses on text and emoji input, no contextual analysis.
Rodrigues et al. [23]	Lisbon Emoji and Emoticon Database (LEED)	Analyzed emoji usage across dimensions such as familiarity, concreteness, and valence in online communication.	Limited to the database's predefined dimensions.
Chandra et al. [24]	Machine learning interpretation of emojis	Developed a system to interpret emojis in sentences, particularly for offensive content, by combining machine learning and deep learning techniques.	Focused on offensive content, limited emoji variety.
Seyednezhad et al. [25]	Emoji usage patterns in short messages	Found that emojis are often used in short messages to express positive sentiments, rather than specific conversation content.	Limited to short messages, lacking deeper analysis.
Ebel et al. [27]	Emoji-only communication	Explored how young people use emoji-only messages, finding that most users increase their understanding of emojis over time.	Focused only on young participants and emoji-only use.
Kelly et al. [19]	Emojis in computer-mediated communication	Emojis help maintain conversations, express feelings, and create an enjoyable interaction environment.	General findings without specific cultural focus.
Gomes et al. [28]	Sentic LDA and sentiment analysis	Enhanced corpus clustering using Sentic LDA, manually tagged based on aspect lexicons; no focus on emojis.	Excluded emojis and emoticons from the analysis.
Khattak et al. [30]	Aspect-based sentiment analysis with LSTM	Used LSTM and hierarchical attention for aspect-based sentiment analysis; excluded emojis and emoticons.	Text-focused, omitted emojis and emoticons.
Rodrigues et al. [31]	Concept-based feature extraction with FCA	Extracted features using FCA for concept-ontology relations; did not consider emojis in data analysis.	Ignored emojis in concept-based analysis.
Koroteey [32]	Co-LSTM for large text sequential analysis	Combined CNN and LSTM for scalable sequential text analysis; excluded emoji data.	Did not include emojis in analysis.
Consoli et al. [33]	Commonsense-based sentiment analysis	Employed ConceptNet and bidirectional LSTM for sentiment analysis, focusing on textual data only.	Did not address the role of emojis.
Ben Ayed et al. [34]	Latent Dirichlet Allocation (LDA) for sentiment	Applied LDA and PLSA for aspect-based sentiment analysis, calculated scores using SimRank algorithm; excluded emojis.	Focused solely on text, excluding emojis.
Liao et al. [35]	Fine-grained aspect-based sentiment (FiGAS)	Focused on sentiment analysis in financial and economic domains, using polarity scores without considering emojis.	Lexicon-based, did not incorporate emojis.

III. PROPOSED SENTIMENT ANALYSIS MODEL

The main problem addressed in this study was to propose an accurate system for reliably sentiment analyzing a message combining text with a string of emojis. The proposed system, illustrated in Fig. 1, comprises two phases: (1) the translation phase and (2) the classification phase.

In the translation phase, the sentiment of the text is analyzed using the Bidirectional Encoder Representations from Transformers (BERT) model, a transformer-based architecture renowned for its ability to capture the contextual meaning of text [36]. BERT processes the input text by considering the entire sentence, allowing it to understand how each word contributes to the overall sentiment. The output from BERT is a contextualized embedding—a vector representation that encapsulates the sentiment conveyed by the text. The vector includes mainly sentiment strength, sentiment polarity, and contextual dependencies. This embedding is then passed through a Long Short-Term Memory (LSTM) network. The LSTM further refines this representation by capturing any sequential dependencies in the sentiment, which is particularly advantageous in understanding how the sentiment might evolve over a longer text. The use of BERT ensures that the model comprehends the nuanced and context-dependent nature of language, while the LSTM helps in maintaining coherence in sentiment analysis over extended sequences.

Simultaneously, the sentiment conveyed by the emojis in the message is interpreted using a different yet complementary approach. Each emoji is first converted into a dense vector representation through an embedding layer using Emoji2Vec [37], which maps emojis into a vector space that reflects their semantic similarities. The word2vec tool transforms words into vectors for computation. This tool defines context based on words that are semantically similar. The skip-gram model in word2vec is preferred over Continuous Bag-of-Words (CBOW) because it predicts context based on current words used. This step is crucial because it translates the often-subtle emotional nuances of emojis into a format that the model can process. The embedded emojis are then processed by an LSTM network, which is particularly effective at capturing the sentiment in sequences of emojis, recognizing how their order and combination might alter the conveyed emotion. The LSTM's

ability to understand sequences makes it particularly well-suited for interpreting the sentiment from emoji sequences, as it can capture changes or continuities in emotional expression across multiple emojis.

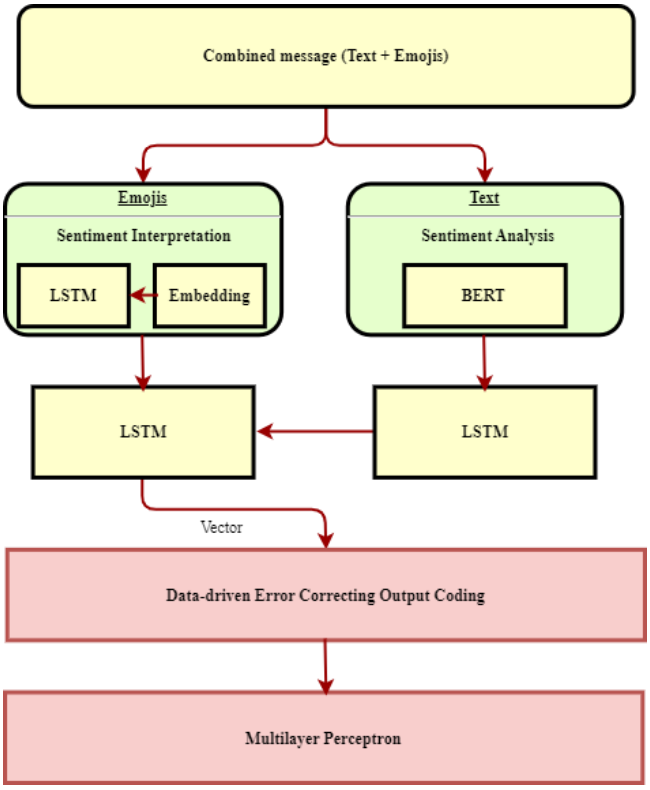


Fig. 1. Sentiment analysis model.

Once the text and emoji sentiments have been individually processed, the outputs from the respective LSTM networks are combined into a single vector that encapsulates the overall sentiment of the message. This combination of text and emoji sentiment ensures that the model considers the full spectrum of emotional expression, leveraging both the explicit and implicit cues provided by the user.

In the classification phase, this combined sentiment vector undergoes further refinement through a data-driven error-correcting output coding layer (DECOC) [38]. The DECOC constructs a coding matrix where each row corresponds to a class and each column to a binary classifier. The classes are encoded as binary strings, allowing the system to decode the output into one of the multiple classes. This approach is particularly effective in handling the variability and complexity of emoji sequences. This layer helps in reducing any noise or errors that may have been introduced during the Translation Phase, thereby enhancing the model's ability to generalize across different types of messages. It ensures that the sentiment analysis remains robust, even in complex scenarios where text and emojis might convey conflicting emotions.

The refined vector is then passed into a Multilayer Perceptron (MLP) which is tasked with classifying the overall sentiment of the message. The MLP's architecture, with its multiple layers of neurons, allows it to perform complex transformations on the input data, gradually refining it until it reaches the final sentiment classification. The use of MLP ensures that the model can accurately classify the sentiment into categories such as positive, negative, or neutral, providing a comprehensive analysis of the emotional content of the message.

Overall, the design leverages advanced NLP techniques, combining the strengths of BERT, LSTM networks, DECOC, and MLPs to offer a robust and accurate sentiment analysis system. The translation phase ensures that the nuanced sentiments in both text and emojis are captured, while the classification phase refines and classifies these sentiments, resulting in a system that can accurately interpret the emotional tone of a combined text and emoji message.

IV. Results and Discussion

The DECOC-MLP classifiers achieved outstanding predictive accuracy across all identified emotional undertones, whether derived from text or emojis, with a high degree of confidence. During the data processing phase, emotional cues are extracted from both the textual and emoji components. The emoji data is processed using a pre-trained model based on the EmojiNet and EmojiString datasets [39], where 80% of the data is randomly assigned for training and 20% for testing. The DECOC classifier generates an output vector populated with potential emotional meanings, while the spaCy library in Python is used to analyze the emotional content of the text.

In the classification phase, the DECOC-MLP model processes keywords and sentences, utilizing a dataset from the Twitter Developer's API [40], which contains 1,440,438 sentences categorized by sentiment. Following training, the model is capable of accurately categorizing words according to

their respective categories. The DECOC-MLP model was trained over 25, 50, 75, and 100 epochs, with batch sizes ranging from 200 to 400. After thorough analysis, an epoch size of 50 and a batch size of 300 were identified as the optimal parameters. Tenfold cross-validation was performed during training, resulting in a mean accuracy of 94.76%, indicating the model's high effectiveness. Accuracy is calculated as the ratio of correctly predicted instances to the total number of instances in the dataset as shown in Eq. (1).

$$Accuracy = \frac{True\ classification}{Total\ of\ samples} \quad (1)$$

The model also demonstrated a low variance value of 2.05×10^{-6}, see Eq. (2), confirming that the predictions are not only accurate but also stable and consistent with expected outcomes.

$$Variance = \frac{\sum(M_i - \overline{M})}{n-1} \quad (2)$$

Where M_i is the value of a specific message, $\overline{M}$ is the mean value, and n is the total number of messages.

The model was validated using the testing set from the Twitter Developer's API dataset and achieved an accuracy of approximately 94%, and a created one based on local survey accomplishing an accuracy of 92.8%, as depicted in Fig. 2.

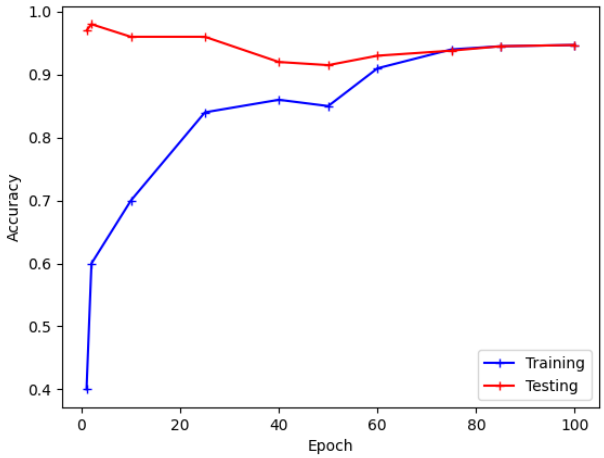


Fig. 2. DECOC-MLP trained-test accuracy.

The Fig. 2 indicates that the model performs well on both the training and testing datasets, but there is a small indication of overfitting. The testing accuracy remains high, which is a positive sign.

To assess the accurate recognition of the emotional undertones of an emoji string in a real case, we designed a survey for Saudis using WhatsApp, consisting of 100 common messages including text and emojis. Table II presents some of these messages.

TABLE II. SAMPLES OF MESSAGES

Message (Text + Emojis)	Translation (Arabic)	Sentiment
"I'm so happy today! ☺️☼ "	"أنا سعيد جدًا اليوم! ☺️☼"	Positive
"I can't believe this happened... 😒💔"	"لا أستطيع تصديق ما حدث... 😒💔"	Negative
"This is just okay. 😐"	"هذا فقط جيد. 😐"	Neutral
"What a wonderful surprise! 🎉"	"يا لها من مفاجأة رائعة! 🎉"	Positive
"I'm really frustrated right now. 😩"	"أنا محبط حقًا الآن. 😩"	Negative
"Feeling a bit down today... ☹️"	"أشعر بالحزن قليلاً اليوم... ☹️"	Negative
"Wow, that was unexpected! 😲"	"واو، لم أتوقع ذلك! 😲"	Neutral to Positive
"Best day ever! 🎉☺️"	"أفضل يوم على الإطلاق! 🎉☺️"	Positive

Therefore, Table III shows a comparison with previous works. The table compares different sentiment analysis approaches over the years, focusing on the data types, classifiers used, datasets, and achieved accuracy. Behera et al. (2021) achieved a high accuracy of 98% using CNN and LSTM on airline reviews based solely on text data. Gupta et al. (2023) incorporated both text and emojis, using multiple classifiers on a large Twitter dataset, reaching an accuracy of 91%. The proposed approach for 2024 also integrates text and emojis, employing advanced techniques like LSTM, DECOC, and MLP, resulting in a strong accuracy of 94% on diverse datasets.

TABLE III. COMPARISON BETWEEN MODELS

Approach	Year	Message	Classifier Used	Dataset	Accuracy
Behera et al. [41]	2021	Text	LSTM CNN	Airline review	98%
Gupta et al. [42]	2023	Emoji + Text	Decision Tree SVM Naïve Bayes	168548 tweets	91%
Proposed approach	2024	Emoji + Text	LSTM DECOC MLP	EmojiString Twitter Developer's API dataset	94%

To sum up, emojis are becoming increasingly important in sentiment analysis as they play a crucial role in conveying emotions in digital communication. This rise in emoji usage does not replace traditional text-based language but rather enhances it as part of the natural evolution of language. As technology continues to influence daily communication, the integration of emojis into sentiment analysis models becomes increasingly significant. The proposed model in this paper focuses on accurately interpreting the emotional undertones conveyed by emoji sequences using DECOC machine learning models. The results demonstrate that the DECOC classifier effectively predicts the sentiment expressed through both textual and emoji inputs with high accuracy.

V. CONCLUSION

This paper highlights the increasing use of emojis on social media platforms among Saudis and presents a machine learning-based model specifically designed for sentiment analysis of messages that combine text and emojis, providing highly accurate emotional assessments. The proposed DECOC-MLP model excels at thoroughly analyzing the emotional content of message components, including both text and emoji sequences, using "EmojiString" and Twitter Developer's API datasets. This dataset allows the model to precisely detect the emotional nuances conveyed by emoji sequences, thereby significantly improving the accuracy of sentiment analysis.

The importance of this approach lies in its potential for future advancements, where the EmojiString dataset can be expanded to include more complex emoji combinations, offering deeper insights into the role of emojis in conveying emotions. This will not only refine the model's accuracy but also contribute to understanding how emojis complement or even replace text in digital communication. Future research should explore the cultural implications of this model, particularly its impact on the evolution of language and communication in various user groups. Recognizing emojis as a fundamental part of language development is essential as they continue to shape the landscape of sentiment analysis.

REFERENCES

[1] K. Garg, "Sentiment analysis of Indian PM's 'Mann Ki Baat'," International Journal of Information Technology, vol. 12, no. 1, pp. 37-48, 2020.

[2] X. Li, B. Wang, L. Li, Z. Gao, Q. Liu, H. Xu, and L. Fang, "Deep2s: improving aspect extraction in opinion mining with deep semantic representation," *IEEE Access*, vol. 8, pp. 104026-104038, 2020.

[3] S. L. Lo, E. Cambria, R. Chiong, and D. Cornforth, "A multilingual semi-supervised approach in deriving Singlish sentic patterns for polarity detection," *Knowledge-Based Systems*, vol. 105, pp. 236-247, 2016.

[4] S. Saha and K. A. Kumar, "Emoji prediction using emerging machine learning classifiers for text-based communication," *J. Math. Sci. Comput.*, vol. 1, pp. 37-43, 2022.

[5] C. I. Eke, A. A. Norman, L. Shuib, and H. F. Nweke, "Sarcasm identification in textual data: systematic review, research challenges and open directions," *Artificial Intelligence Review*, vol. 53, pp. 4215-4258, 2020.

[6] S. Zobeidi, M. Naderan, and S. E. Alavi, "Opinion mining in Persian language using a hybrid feature extraction approach based on convolutional neural network," *Multimedia Tools and Applications*, vol. 78, no. 22, pp. 32357-32378, 2019.

[7] S. S. Mujawar and P. R. Bhaladhare, "An aspect based multi-label sentiment analysis using improved BERT system," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, pp. 228-235, 2023.

[8] N. Rethmeier and I. Augenstein, "A primer on contrastive pretraining in language processing: Methods, lessons learned, and perspectives," *ACM Computing Surveys*, vol. 55, no. 10, pp. 1-17, 2023.

[9] L.-C. Chen, K.-H. Chang, S.-C. Yang, and S.-C. Chen, "A corpus-based word classification method for detecting difficulty level of English proficiency tests," *Applied Sciences*, vol. 13, no. 3, p. 1699, 2023.

[10] S. Gupta, A. Singh, and J. Ranjan, "Multimodal, multiview and multitasking depression detection framework endorsed with auxiliary sentiment polarity and emotion detection," *International Journal of System Assurance Engineering and Management*, vol. 14, no. Suppl 1, pp. 337-352, 2023.

[11] Z. Wei, Y. Chen, Q. Zhao, P. Zhang, L. Zhou, J. Ren, Y. Piao, B. Qiu, X. Xie, S. Wang, et al., "Implicit Perception of Differences between NLP-Produced and Human-Produced Language in the Mentalizing Network (Adv. Sci. 12/2023)," *Advanced Science*, vol. 10, no. 12, 2023.

[12] E. Cambria and A. Hussain, "Sentic applications," *Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis*, pp. 107-153, 2015.

[13] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, "Affective computing and sentiment analysis," *A Practical Guide to Sentiment Analysis*, pp. 1-10, 2017.

[14] M. S. Mahmud, J. Z. Huang, S. Salloum, T. Z. Emara, and K. Sadatdiynov, "A survey of data partitioning and sampling methods to support big data analysis," *Big Data Mining and Analytics*, vol. 3, no. 2, pp. 85-101, 2020.

[15] S. Al Rashdi, "Functions of emojis in WhatsApp interaction among Omanis," *Discourse, Context & Media*, vol. 26, pp. 117-126, 2018.

[16] A. Albawardi, "The translingual digital practices of Saudi females on WhatsApp," *Discourse, Context & Media*, vol. 25, pp. 68-77, 2018.

[17] R. Kelly and L. Watts, "Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships," in *Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances, and Design*, 2015.

[18] S. Al-Azani and E.-S. M. El-Alfy, "Combining emojis with Arabic textual features for sentiment classification," in *2018 9th International Conference on Information and Communication Systems (ICICS)*, 2018, pp. 139-144.

[19] E. Dresner and S. C. Herring, "Functions of the nonverbal in CMC: Emoticons and illocutionary force," *Communication Theory*, vol. 20, no. 3, pp. 249-268, 2010.

[20] L. A. G. Neel, J. G. McKechnie, C. M. Robus, and C. J. Hand, "Emoji alter the perception of emotion in affectively neutral text messages," *Journal of Nonverbal Behavior*, vol. 47, no. 1, pp. 83-97, 2023.

[21] S. Al Rashdi, 2015. Forms and functions of emojis in WhatsApp interaction among Omanis (Doctoral dissertation, Georgetown University).

[22] I. Ebel and J. P. Dutra, "Do You Speak Emoji? The Language of Emojis," *The SAGE Handbook of Social Media Marketing*, p. 320, 2022.

[23] S. M. M. Seyednezhad, H. Fede, I. Herrera, and R. Menezes, "Emoji-word network analysis: Sentiments and semantics," in *The Thirty-First International FLAIRS Conference*, 2018.

[24] Y. Ma, H. Peng, and E. Cambria, "Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[25] P. Chandra and U. Prasad, "Classification of Emojis using Artificial Neural Network and Natural Language Processing," in *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*, 2021, pp. 205-212.

[26] M. Dragoni, S. Poria, and E. Cambria, "OntoSenticNet: A commonsense ontology for sentiment analysis," *IEEE Intelligent Systems*, vol. 33, no. 3, pp. 77-85, 2018.

[27] A. V. Yeole, P. V. Chavan, and M. C. Nikose, "Opinion mining for emotions determination," in *2015 International Conference on*

Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015, pp. 1-5.

[28] R. F. Gomes and B. Casais, "Feelings generated by threat appeals in social marketing: text and emoji analysis of user reactions to anorexia nervosa campaigns in social media," *International Review on Public and Nonprofit Marketing*, vol. 15, no. 4, pp. 591-607, 2018.

[29] S. Poria, I. Chaturvedi, E. Cambria, and F. Bisio, "Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis," in *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 4465-4473.

[30] A. Khattak, M. Z. Asghar, Z. Ishaq, W. H. Bangyal, and I. A. Hameed, "Enhanced concept-level sentiment analysis system with expanded ontological relations for efficient classification of user reviews," *Egyptian Informatics Journal*, vol. 22, no. 4, pp. 455-471, 2021.

[31] D. Rodrigues, M. Prada, R. Gaspar, M. V. Garrido, and D. Lopes, "Lisbon Emoji and Emoticon Database (LEED): Norms for emoji and emoticons in seven evaluative dimensions," *Behavior Research Methods*, vol. 50, pp. 392-405, 2018.

[32] M. V. Koroteev, "BERT: a review of applications in natural language processing and understanding," *arXiv preprint arXiv:2103.11943*, 2021.

[33] S. Consoli, L. Barbaglia, and S. Manzan, "Fine-grained, aspect-based sentiment analysis on economic and financial lexicon," *Knowledge-Based Systems*, vol. 247, p. 108781, 2022.

[34] M. B. Ayed, F. Bouchhima, and M. Abid, "Automated fingerprint recognition using the decoc classifier," *International Journal of Computer Information Systems and Industrial Management Applications*, 2012, pp. 2150-7988.

[35] J. Liao, M. Wang, X. Chen, S. Wang, and K. Zhang, "Dynamic commonsense knowledge fused method for Chinese implicit sentiment analysis," *Information Processing & Management*, vol. 59, no. 3, p. 102934, 2022.

[36] B. Eisner, T. Rocktäschel, I. Augenstein, M. Bošnjak, and S. Riedel, "emoji2vec: Learning emoji representations from their description," *arXiv preprint arXiv:1609.08359*, 2016.

[37] A. Pradhan, M. R. Senapati, and P. K. Sahu, "Improving sentiment analysis with learning concepts from concept, patterns lexicons and negations," *Ain Shams Engineering Journal*, vol. 13, no. 2, p. 101559, 2022.

[38] R. K. Behera, M. Jena, S. K. Rath, and S. Misra, "Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data," *Information Processing & Management*, vol. 58, no. 1, p. 102435, 2021.

[39] M. B. Ayed and A. Alsaawi, "A Novel Machine Learning Model for Predicting the Meaning of an Emojis String in Social Media Platforms," Studies in Informatics and Control, vol. 33, pp. 91–98, 2024.

[40] K. Chouhan, M. Yadav, R. K. Rout, K. S. Sahoo, N. Z. Jhanjhi, M. Masud, and S. Aljahdali, "Sentiment Analysis with Tweets Behaviour in Twitter Streaming API," *Comput. Syst. Sci. Eng.*, vol. 45, no. 2, pp. 1113-1128, 2023.

[41] R. K. Behera, M. Jena, S. K. Rath, and S. Misra, "Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data," *Information Processing & Management*, vol. 58, no. 1, p. 102435, 2021.

[42] S. Gupta, A. Singh, and V. Kumar, "Emoji, text, and sentiment polarity detection using natural language processing," *Information*, vol. 14, no. 4, p. 222, 2023.

A Smart IoT System for Enhancing Safety in School Bus Transportation

Yousef H. Alfaifi[1], Tareq Alhmiedat[2]*, Emad Alharbi[3], Ahad Awadh Al Grais[4],
Maha Altalk[5], Abdelrahman Osman Elfaki[6]*
Faculty of Computers and Information Technology, University of Tabuk, Tabuk, 71491, Saudi Arabia[1, 2, 3, 6]
Artificial Intelligence and Sensing Technologies (AIST) Research Center, University of Tabuk, Tabuk, 71491, Saudi Arabia[2]
SEHA Virtual Hospital, Ministry of Health, Digital City, Riyadh, Saudi Arabia[4, 5]

***Abstract*—School districts globally implement comprehensive and expensive strategies to offer safe bus transportation to and from school. However, these technologies are unfeasible to schools with limited financial resources, thereby leaving students at risk of serious injury. This study focuses on five major obstacles to safe school transportation: 1) forgetting students on the bus unattended; 2) students' abnormal behavior; 3) overcrowding; 4) abnormal driver behavior; and 5) the risk of a bus running over children after they have disembarked. This paper developed an intelligent system using the Internet of Things, including rule-based and mathematical solutions to overcome the five transportation safety issues in student buses mentioned above, and to enhance student safety, using a bracelet system, short- and long-RFID sensors, and a processing unit to monitor the bus and its surrounding area. Therefore, in this paper, the proposed solution is superior to previous works in the same field. It is distinguished by its comprehensiveness and reasonable cost, making it affordable and easy to both install and maintain.**

***Keywords*—IoT; safety; RFID; school bus; transportation**

I. INTRODUCTION

Keeping students safe on their journey to and from school is of great importance not only to parents, school officials, and administration, but to the entire local community. School buses play a crucial role in transporting students to and from school, serving as a prevalent and cost-effective mode of transportation in many educational communities and countries. Public institutions, communities, and nations aim to expand and improve their school bus services to lower transportation costs, alleviate traffic congestion, and decrease carbon emissions.

Despite these benefits, ensuring safe transportation for students can be difficult due to five safety issues that might harm their safety. The first is when children are forgotten and left on the bus. It is difficult to keep track of all the children to make sure everyone gets off the bus. Left unattended for long periods of time, these forgotten children can harm their health or even cause death [1, 2]. Second, student fights on school transport can hinder their and others' safety, as can other abnormal behavior on the school bus [3, 4]. Some parents complain about fights on the school bus as their children exhibit signs of aggression. It is challenging for the transportation supervisor or driver to monitor the children throughout the entire trip, particularly on a crowded bus.

The third issue is that overcrowded buses where that pose significant risks to students' safety and well-being, as it can lead to discomfort, accidents, delays, and difficulties in emergency evacuations, all of which put students at risk [5]. The fourth impediment to student safety is the unruly behavior of drivers [6]. Such behavior can create a negative environment on the bus, leading to distractions, discomfort, and potentially unsafe conditions for the children. Finally, the fifth obstacle to student safety is accidents that occur when the children disembark. As they cross the road or move away from the bus, they may unknowingly position themselves in the bus's blind spots, outside of the driver's view. If the bus starts moving, there is a risk of the child being injured or killed [7, 8].

Recent technological advancements, especially within the Internet of Things (IoT), play a crucial and effective role in providing safe and effective solutions to various problems [9]. IoT refers to the network of physical objects, or "things," embedded with technologies, software, and sensors, which exchange data with other devices and systems over the internet [10]. This paper applies IoT solutions to handle the five safety issues and enhance security in school transportation. To prevent children from being forgotten on the bus, this paper suggests the use of a bracelet system, which is scanned when the student enters the bus and again when they exit. This solution tracks the boarding and disembarking times and locations, as well as the duration of the time on the bus.

The bracelets, which would be equipped with location-tracking capabilities, should transmit location data to a central system capable of receiving and processing this data. This location data could solve multiple safety issues. In addition to monitoring students who are forgotten on the bus, it could also be used to identify abnormal student behavior via a special algorithm (criteria for a fight could be included frequent position changes, the simultaneous swapping of places by multiple students, or a large gathering of students in one area); location data could also trigger automated alerts when buses surpass safe occupancy levels. In addition, the bracelet could also generate alerts for abnormal driver behavior using the location data. Finally, location data could be used to notify drivers if a student or students are in the bus's blind spot; the driver would then be restricted from moving until all students are a safe distance from the bus.

The contribution of our proposed system is to provide a comprehensive set of safety functionalities at a very minimal cost and with low computational power requirements. This makes our proposed system accessible and applicable to a wide

*Corresponding Author.

range of users. The remaining sections of the paper are organized as follows: Section II discusses recent and updated related works. Section III describes the mathematical model used to develop this smart IoT system. Section IV contains a detailed explanation of the system's implementation. Section V presents a validation of the proposed IoT system. Finally, the discussion and conclusion are presented in Section VI.

II. Related Work

This section discusses related research on IoT systems and safety. The timeframe was limited to the period from 2019 to 2024, as technology has been evolving rapidly, and any work published before this period could be considered outdated. Two factors served as inclusion criteria. Firstly, the work must address one or more of the problems outlined in the introduction (Section I). Secondly, the work should offer practical, and implementable, solutions. All theoretical works were excluded, as well as any redundant ideas.

Abbas et al., [11] addressed the issue of students being inadvertently left on school buses by introducing a system that enabled parents to monitor their children's daily commute using GPS technology. This system alerted parents before the bus reached their home, ensuring the safety of their children. In study [12], Liu and Peng, addressed safety incidents where young children were left unattended on school buses during the summer by implementing a Force Sensing Resistor (FSR) sensor array. This system detected the movement of a forgotten child based on the resistance of the FSR module. Krishnan and Vasuki, in study [13], created an intelligent system employing Global System for Mobile applications (GSM) and Radio Frequency Identification (RFID) technologies. RFID was utilized to recognize students and link them to their parent's contact information, with GSM serving as the medium to notify parents of their child's whereabouts via text message. Paul and Sasirekha, in the study [14], developed a smart system by utilizing three technologies—global positioning system (GPS), RFID and GSM—to ensure a safe school bus journey. RFID oversaw children entering and exiting the bus and sent SMS notifications to parents. If the child didn't reach school or home on time, parents could activate the registered number on the GPS tracker. The tracker then monitored the child's position, providing the latitude and longitude coordinates to the parent's mobile device.

Mott and Kotla, in [15], addressed the issue of children being run over after disembarking from the school bus by developing a detection system based on a LiDAR unit and a dashcam, both controlled by a Raspberry Pi computer. Kumari et al., in [16], created an IoT- and Android-based system that allowed parents and schools to monitor the comfort and safety conditions on the bus in real-time. Gadekar et al., in [17], developed a smart school bus tracking system using QR codes to monitor student boarding and disembarking. To track bus movement, they created a mobile application utilizing the Django framework and GPS integrated with the Google Maps API, enabling parents to stay informed about the real-time location of the bus. In Naik et al.'s research [18], a school bus monitoring system was developed to share details on the bus's location, speed, estimated arrival time, and the students' whereabouts. This functionality was facilitated by IoT technology. The tracking of the school bus was carried out using GPS technology. Parents are alerted through GSM. Everyone is identified by a unique ID through RFID. The bus tracking system allowed for real-time tracking on smartphones. Additionally, the researchers incorporated an algorithm to track the bus and estimate its arrival time.

Sakphrom et al., in [19], developed a smart system for monitoring children inside a school bus by utilizing Bluetooth Low Energy (BLE), which is compatible with a signal strength indicator (RSSI) algorithm. The system monitored children entering and exiting the bus and notified the driver to help them ascertain whether any child was left on-board. The system assessed the distance between devices to determine the current position of the children. To achieve this, this study employed a simplified and highly accurate machine learning approach known as the least mean square (LMS) algorithm, combined with model-based RSSI localization techniques. Malathy et al., in [20], introduced a smart module that relied on GPS to transmit data to a remote server via Wi-Fi. The data uploaded was subsequently accessible to the client through a mobile application, which retrieved the information and mapped the vehicle's location. An alert system activated the microcontroller to generate a push notification from the server script upon scanning the student's RFID tag with the RFID reader, signaling to the relevant authorities and parents that the student has boarded the bus.

Wang et al., [21], developed an abnormal behavior prediction method by integrating multiple indices of student behavior and text information within a big data environment. They utilized an optimized K-means algorithm for clustering student behavior data and combined it with long-short-term memory (LSTM) networks and neural factorization machines for predictive analysis. This study enhanced the field's ability to identify and manage potentially dangerous behaviors in educational settings. Ding et al., [22], developed an intelligent system to detect abnormal behavior in students using human skeleton analysis and deep learning. They utilized the OpenPose deep learning network to extract spatiotemporal features of the human skeleton and classified these features with a graph convolutional neural network to reduce computational complexity. The system demonstrated high accuracy, with a recognition rate exceeding 99.50%. These results underscore the system's potential for real-time monitoring of student behavior on school buses, enhancing safety measures through effective detection and classification of abnormal behavior. Murawski et al., [23], deployed an intelligent system for overseeing school bus drivers using a "stability" index, which contrasts with the 11 risky driving behaviors determined by the digital tachograph (DTG) and GPS integrated with the Google Maps API.

Etaati et al., [24], evaluated the effectiveness of promoting active transportation to school by utilizing the Safe Routes to School (SRTS) program to estimate the likelihood of students choosing to walk to school. They employed SHapley Additive exPlanations to identify significant variables that influenced active transportation behavior in their models. The findings highlighted factors such as home-to-school distance and safety concerns, including crime rates and traffic speed along the route. In this solution, however, they replaced the bus transportation system with a safe road, which may not be applicable in all locations.

TABLE I. SUMMARY OF RELATED WORKS

No	Work	Techniques/ Algorithms	Limitation
1	Krishnan and Vasuki, in [13]	SMS and GSM	The speed and cost are affected by the use of GSM technology.
2	Abbas et al., in [11]	GPS and Mobile application	There is no monitoring system in place for the students themselves; instead, manual checks are required to ensure students board the bus.
3	Liu and Peng, in [12]	Force Sensing Resistor and android application	The accuracy could be affected by the movements of students.
4	Paul and Sasirekha, in [14]	GPS, RFID and GSM	It is very difficult to provide real-time data.
5	Mott and Kotla, in [15]	LiDAR unit a dashcam, both, and Raspberry Pi	High processing cost.
6	Kumari et al., in [16]	Dashcam, Android	High-cost devices are challenging to provide.
7	Gadekar et al., in [17]	QR Code, Google Maps API, Django	Scanning the QR code of every student when boarding and disembarking slows the system and makes it impractical.
8	Naik et al., in [18]	GPS, GSM, RFID, and Android application	High-cost devices are challenging to provide to everyone.
9	Sakphrom et al., in [19]	Bluetooth Low Energy (BLE), signal strength indication (RSSI) algorithm, and least mean square (LMS) algorithm	High processing cost.
10	Malathy et al., in [20]	GPS, Wi-Fi, RFID, and Android application	High processing cost.
11	Ding et al., in [22]	OpenPose deep learning network Graph convolutional neural network (GCN) Sliding window voting method	High processing cost.
12	Wang et al., in [21]	Optimized K-means, Long-Short-Term Memory (LSTM) networks, Neural Factorization Machines (NFM)	High processing cost.
13	Murawski et al., in [23]	DTG and GPS integrated with the Google Maps API	High-cost devices are challenging to provide to everyone.
14	Etaati et al., in [24]	random forest, logistic regression, and support vector machines	The researchers replaced the bus transportation system with a safe road, which may not be applicable in all locations.
15	LR et al., in [25]	RFID and GPS	Does not provide complete solutions.
16	Gadade et al., in [26]	ESP32 microcontroller, GPS module, RFID reader, RTC module, LCD	High-cost devices are challenging to provide to everyone.
17	Kumari et al., in [27]	Python-configured PC with a camera and OpenCV	High processing cost.

LR et al., in [25], developed a system that addressed safety standards by ensuring continuous student monitoring through RFID and GPS technologies. This system recorded the boarding and disembarking times of students passing through two checkpoints, tracked the real-time location of the bus, and monitored its speed to comply with regulations. In Gadade et al.'s research [26], a smart school bus and student tracking system based on IoT was developed utilizing an ESP32 microcontroller, GPS module, RFID reader, RTC module, LCD display, and integration with Google Sheets. The ESP32 microcontroller functioned as the central processing unit. The GPS module facilitated the tracking of the school bus's location, enabling real-time monitoring for parents, school staff, and authorities.

Furthermore, RFID technology was employed to monitor students during boarding and disembarking. For accurate timekeeping, a Real-Time Clock (RTC) module was integrated to synchronize system time and ensure precise event logging. Additionally, an LCD display served as a local interface, offering real-time updates on bus location, current time, and important announcements. Kumari et al., in [27], created a smart system utilizing a Python-configured PC with a camera and OpenCV for instant facial recognition of students during school bus boarding and exiting. Upon identifying a student, the system automatically sent an email notification to inform parents, providing reassurance to caregivers. Alongside student recognition, the system included alcohol detection features for the driver. In cases where the driver was deemed intoxicated, the bus's ignition system would lock, preventing unsafe driving.

Table I, illustrates this analysis of the related works, defining the techniques or algorithms that each utilized and detailing their limitations. The related works have been analyzed by defining the strengths and weaknesses of each work. The datasets are associated with their related work. Similar works have discussed safety in school transportation systems. The database used in the work is specific to each study, and currently, there is no general-use database available.

III. MODELING

This section presents a mathematical model that serves as the foundation for implementing this paper's five solutions. There are three assumptions, which are:

Assumption 1: Each student should wear a bracelet.

Assumption 2: The bus driver should wear a bracelet.

Assumption 3: Each bracelet has a unique integer number.

Fig. 1 displays the variables utilized in these proposed solutions, which are incorporated into the mathematical model.

1. s → student
2. b → bracelet
3. seat → bus seat
4. r1 → first reader
5. r2 → second reader
6. r3 → inside reader
7. r4 → inside reader
8. r5 → inside reader
9. bdrv → driver bracelet
10. driver_set → driver seat

Fig. 1. Variables of the proposed solutions.

In the following formulas, the mathematical model has been described in form on First Order Logic (FOL), which is suitable for describing real life problems [28].

$$\forall b: bracelet\ (b) \wedge integer(number) \implies unique(b, number) \quad (1)$$

Rule 1 denotes that there is a unique integer number assigned to each bracelet, which functions as the primary key or identifier for a bracelet.

$$\exists s\ \exists b: student(si) \wedge bracelet\ (bj) \wedge wear\ (si\ , bj) \implies assign(si, bj) \quad (2)$$

Rule 2 denotes that each student has their own unique bracelet, which means the reading of a specific bracelet number identifies the associated specific student.

$$\forall s\ \forall r1\ \forall r2\ \forall ti\ \forall b: reader(r1) \wedge reader(r2) \wedge time(t) \wedge bracelet\ (bj) \wedge assign(si, bj) \wedge read(r1, bi, t1) \wedge read(r2, bi, t2) \wedge (t2 > t1) \implies si\ boarding\ the\ bus \quad (3)$$

Rule 3 denotes that if the bracelet of student (i) was read by reader 1 first and then read by reader 2 later, this means that student (i) has boarded the bus.

$$\forall s\ \forall r1\ \forall r2\ \forall ti\ \ \forall b: reader(r1) \wedge reader(r2) \wedge time(t) \wedge bracelet\ (bj) \wedge assign(si, bj) \wedge read(r1, bi, t1) \wedge read(r2, bi, t2) \wedge (t1 > t2) \implies si\ gets\ off\ the\ bus \quad (4)$$

Rule 4 denotes that if student (i)'s bracelet was read by reader 2 first and then reader 1 later, then the student (i) has gotten off the bus.

$$\forall s: true\ (rule\ 3) \wedge true(rule\ 4) \implies time(si, boarding\ and\ alighting) \wedge location(si) \quad (5)$$

Rule 5 indicates that once rules 3 and 4 are met, the boarding and disembarking of a student have been established, along with their location during the bus ride. Rules 3, 4, and 5 present the mathematical formulation for the first solution, which confirms that no students were forgotten on the bus.

$$\forall bn\ \forall ri\ \forall seatj\ \forall sx: ((ri == reader\ 3) \wedge (ri == reader\ 4) \wedge (ri == reader\ 5)) \wedge student(sx) \wedge seat(seatj) \wedge set(setj, sx) \wedge wear\ (si\ , bn) \wedge read(ri, bn) \implies assign(sx\ , seat\ j) \wedge location(seatj) \quad (6)$$

Rule 6 specifies that the three readers, r1, r2, and r3, are used to determine the student's seat. After the student takes a seat in a specific location, the three readers employ the triangulation method to identify the student's seat and exact location.

$$\forall ri\ \forall bx\ \forall setj\ \forall sx: ((ri == reader\ 3) \wedge (ri == reader\ 4) \wedge (ri == reader\ 5)) \wedge seat\ (setj) \wedge bracelet\ (bx) \wedge assign(setj, sx) \wedge ((distance(ri, bx) == n) \wedge ((distance(ri, setj) == m) \wedge (n \neq m) \implies sx\ in\ incorrect\ seat \implies send(alarm) \quad (7)$$

Rule 7 states that if the distance between a particular reader i and a specific seat j (predefined distance) assigned to student x is not equal to the distance between bracelet x assigned to student x and the particular reader i, this indicates that student x is seated in the incorrect seat; an alarm will be sent to administrators.

$$\forall ri\ \forall b\ \forall setj\ \forall s: ((ri == reader\ 3) \wedge (ri == reader\ 4) \wedge (ri == reader\ 5)) \wedge bracelet\ (bx) \wedge bracelet(by) \wedge (distance(ri, bx) == n) \wedge (distance(ri, by) == m) \wedge (m \wedge n) < safe\_distance \implies student\_gathering \implies send(alarm) \quad (8)$$

Rule 8 states that if the distance between two or more students is less than the predefined safe distance that should be maintained, indicating a group gathering, an alarm will be triggered and sent to the administrators as an alert. Rules 6, 7, and 8 illustrate the mathematical representation for the second solution, which seeks to prevent student fights and other erratic behavior.

$$\forall r1\ \forall r2\ \forall bi: reader(r1) \wedge reader(r2) \wedge bracelet\ (bi) \wedge read(r1, bi) \wedge \neg read(r2, bi) \wedge count(bi, n) \wedge (n > bus\_capacity) \implies send(alarm) \quad (9)$$

Rule 9 states that if the number of students who have been scanned by the first reader but not by the second reader exceeds the bus capacity, the system will send an alert to the administrators to indicate overcrowding. Rule 9 illustrates the mathematical representation for the third solution.

$$\exists bdrv\ \exists r1\ \exists driver\_set: reader(r3) \wedge reader(r4) \wedge reader(r5) \wedge bracelet\_driver\ (bdrv) \wedge (distance(ri, driver\_set) == x) \wedge$$

$$distance(r1, bdrv) > x \implies driver\ misplaced \wedge send(alarm) \quad (10)$$

Rule 10 denotes that if the distance between the driver's bracelet and the readers is greater than the distance between the first reader and the driver's seat, it indicates that the driver is not in the correct position and triggers an alarm to administrators. Rule 10 presents the mathematical formulation for the fourth solution, which seeks to prevent erratic driver behavior.

$$\forall ri\ \forall b\ \forall s: ((ri == reader\ 1) \wedge (ri == reader\ 2) \wedge (ri == reader\ 3) \wedge (ri == reader\ 4) \wedge (ri == reader\ 5)) \wedge bracelet(bx) \wedge assign(sx, bx) \wedge out\_range(ri, bx) \implies bus\ save\ to\ move \quad (11)$$

Rule 11 mandates that the bus will only move when all students are out of the reader's range, ensuring that the bus does not depart while students are still within proximity. This rule introduces the mathematical formulation for the fifth solution, which addresses student safety after they've disembarked from the bus. Table II, displays the rules and their corresponding solutions.

TABLE II. THE RULES AND THEIR CORRESPONDING SOLUTIONS

Rule	Solution
3, 4, and 5	First solution, which prevents students being forgotten on the bus: track the boarding and disembarking times, locations, and the duration of the student's time inside the bus.
6, 7, and 8	Second solution, which prevents erratic student behavior: set thresholds for the frequency of place changes, the number of students swapping places simultaneously, and instances of groups of students gathering in one place.
9	Third solution, which prevents bus overcrowding: trigger automated alerts for administrators when buses surpass safe occupancy levels.
10	Fourth solution, which prevents erratic driver behavior: track driver movement and generate alerts for abnormal driving behavior, such as frequent changes in location
11	Fifth solution, which prevents students from being injured or killed when in the bus's blind spot: prevent the bus from moving if students are still in the bus area. The driver is restricted from moving until all students are safely outside the bus area.

IV. IMPLEMENTATION

This section explains the implementation of the aforementioned rules in order to achieve increased bus safety. The implementation system has been developed based on different components. The main concept of the developed smart IoT system for enhancing safety in school bus transportation is presented in Fig. 2.

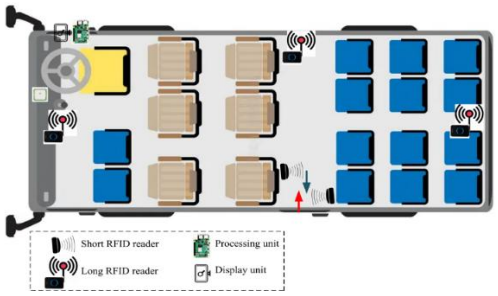


Fig. 2. The main components of the developed smart IoT system.

The entrance of the bus employs two short-range RFID readers, as presented in Fig. 2. There are two main reasons for utilizing these readers: the first is to identify students as they enter or leave the bus, and the second is to determine the direction of the student (entering or leaving). For instance, if reader A detects the RFID signal of the student before reader B does, then the student is entering the bus; whereas if reader B detects the RFID signal of the student first, then the student is leaving the bus. Algorithm 1 presents the method for determining the student's direction (entering or exiting the bus). This ensures that students are not forgotten on the bus.

Algorithm 1: Determine the status of the student (enter / exit the bus)

```
let sb_i is the student's bracelet of student i
let rf_e is the first short-range RFID reader
let rf_x is the second short-range RFID reader
let status is the current status of the student (1: enter and 2: exit)
if rf_e detects sb_i then
    if rf_x detects sb_i then
        return 1
else if rf_x detects sb_i then
    if rf_e detects sb_i then
        return 2
```

One of the main benefits offered by the developed IoT system is its continuous monitoring of children's positioning on the bus. Three long-range RFID readers are employed to triangulate the children's locations. When a child enters the bus, the system records the location for each child and stores these locations in a local database. Subsequently, the system continuously estimates the location of each child to keep track of their initial location. The code used to estimate the continuous locations of the students and check for changes in position is shown in Algorithm 2.

Algorithm 2: Check for students change-position in the bus

```
let sb_i is the student's bracelet of student i
let student[] is the array of students in the bus
let init_pos[] is the array of 2D positions for students in the bus
let new_pos[] is the array of the new estimated 2D positions for students in the bus
let change-position = False
while (every 20 seconds):
    for each student in new_pos[]:
        if dist(init_pos[], new_pos[]) >= 100 cm then
            change-position = True
            show alarm in driver display
            send alarm to administrators
End
```

When one or more students have changed their initial locations (seats), the driver will be informed by displaying an alarm on the screen, and an alarm will also be sent to the administrators. Furthermore, the developed IoT system can detect crowds, such as more than two students in the same position on the bus. In such cases, the driver will be informed by

displaying an alarm on the screen, and an alert will be sent to the administrators. Algorithm 3 is utilized to check for any gatherings (crowds) inside the bus. Algorithms 2 and 3 are utilized to implement the second solution, which involves detecting abnormal behavior inside the bus.

Algorithm 3: Check for students gathering inside the bus

0: let sb_i is the student's bracelet of student *i*
1: let *student[]* is the array of students in the bus
2: let *init_pos[]* is the array of 2D positions for students in the bus
3: let *new_pos[]* is the array of the new estimated 2D positions for students in the bus
4: let *gathering* = False
5: **while** (every 20 seconds):
6: **for each** *position* in *new_pos[]*:
7: **if any dist**(*position*, *position*+1, *position*+2) <= 50 cm **then**
9: *gathering* = True
10: show alarm in driver display
11: send alarm to administrators
12: **End**

The long-range RFID readers are utilized to perform a significant localization task. For example, when the system detects a person leaving the bus, the long-range RFID readers begin tracking the location of that student. If the RFID readers no longer receive signals from the student's bracelet, the driver is alerted to safely depart from the drop-off point. However, if the student remains within range of one or more long-range RFID readers, the driver is instructed to wait until the student moves away from the bus. Algorithm 4 outlines the method used to estimate the distance between the child and the bus, to ensure that children standing in the bus's blind spot are not injured or killed.

Algorithm 4: Check for student's distance to bus

0: let sb_i is the student's bracelet of student *i*
1: let lRF_1 is the first long-range RFID reader
2: let lRF_2 is the second long-range RFID reader
3: let lRF_3 is the third long-range RFID reader
4: let *must_stop* = True
4: **if** status == 2: (Algorithm 1)
5: **if** sb_i in the range of (lRF_1 or lRF_2 or lRF_3) **then**
6: *must_stop* = True
7: display alarm
8: **else if** sb_i **not** in the range of (lRF_1 or lRF_2 or lRF_3) **then**
9: *must_stop* = False
10: **End**

Moreover, the developed system can monitor the driver's behavior by continuously tracking the driver's position. Each driver possesses a unique bracelet for identification purposes. Consequently, when the driver relocates from their initial position, the coordinator is promptly notified to prevent accidents and increase safety. Algorithm 5 demonstrates this method.

Algorithm 5: Determine the location of the driver

0: let db_i is the student's bracelet of student *j*
1: let $init\_loc_{dj}$ is the initial location of the driver *j*
2: let lRF_1 is the first long-range RFID reader
3: let lRF_2 is the second long-range RFID reader
4: let lRF_3 is the third long-range RFID reader
5: estimate_driver_loc(lRF_1, lRF_2, lRF_3)
6: let cur_loc_{dj} is the current location of the driver *j*
7: **while** (True):
8: **if** ($init\_loc_{dj}$!= cur_loc_{dj}) **then**
9: emergency_call()
10: send alarm
11: **else:** cur_loc_{dj} = estimate_driver_loc(lRF_1, lRF_2, lRF_3)
12: **End**

Furthermore, the system has the capability to calculate the total students aboard the bus to prevent overcrowding issues. This count is determined by the short-range RFID readers located at the bus entrance. If the number of students onboard exceeds the bus capacity, the coordinator is promptly alerted. Algorithm 6 demonstrates this, below.

Algorithm 6: Count the total number of students inside a bus

0: let sb_i is the student's bracelet of student *i*
1: let rf_e is the first short-range RFID reader
2: let rf_x is the second short-range RFID reader
3: let cap_k is the capacity of the bus *k*
4: let $count_k$ is the total number of students inside the bus *k*
5: **if** ($count_k$< cap_k) **then**
6: emergency_call()
7: send alarm
8: **End**

The developed system was implemented using different components (processing, RFID readers, and communication), as follows:

1) Processing unit: Raspberry Pi 4 processes the data received from short- and long-range RFID readers, and takes suitable action accordingly.

2) Long-range RFID readers: These are required to obtain the precise position of the student on the bus. Three long-range RFID readers were deployed in the bus area in order to obtain accurate localization information from each child.

3) Short-range RFID readers: Two short-range RFID readers were employed at the bus door to identify the direction of the student (entering or leaving the bus).

4) Communication unit: For reliable communication between the above units, processed data, and short- and long-range RFID readers, this paper utilized WiFi (IEEE 802.11) network. Fig. 3 presents the communication process between the above components; it also shows the proposed IoT system architecture. Fig. 4 shows the communication process between the main components.

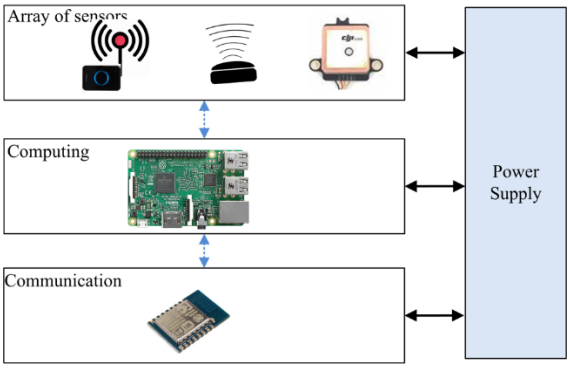


Fig. 3. The proposed IoT system architecture.

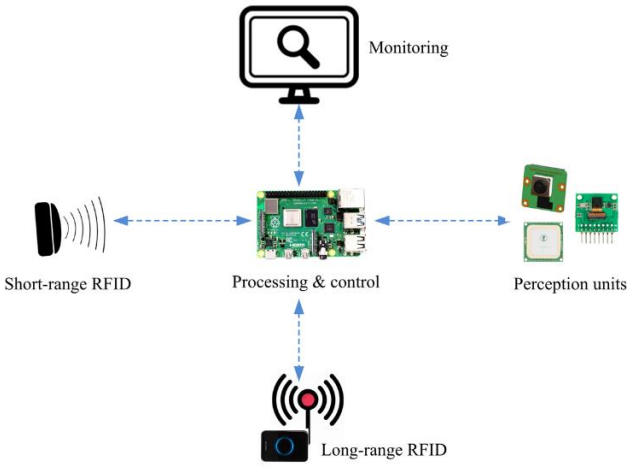


Fig. 4. The communication process between the main components.

V. VALIDATION

This paper's solutions were tested using simulation software designed by implementing the aforementioned algorithms to replicate various threats to students' safety. In an IoT environment, it is an acceptable practice to use simulated scenarios to test and demonstrate the accuracy and scalability of the proposed solutions [29, 30].

A. First Scenario

A bus with 30 seats (s1, s2, …, s30) has 30 students (st1, st2, …, st30) entering the bus and sitting randomly in the seats. The bus has been simulated as an array, and the students are represented as variables occupying the array positions. Additionally, the researchers entered an excess number of students to exceed the bus capacity in order to test rule 9.

To verify the accuracy of the IoT system, this paper compared the data from the readers' database with the array content, achieving a consistent 100% accuracy. The experiment was replicated with buses having 45 seats and 45 students, 60 seats and 60 students, 75 seats and 75 students, and 90 seats and 90 students. Assuming a one-hour bus journey, readings were recorded every 30 seconds, totaling 120 intervals. Accuracy remained at 100% throughout.

B. Second Scenario

In the same simulation environment as before, a bus with 31 seats (s1, s2, …, s30) and 30 students (st1, st2, …, st30) was considered, with the first seat reserved for the driver. Therefore, the first position of the array is occupied by a special variable 'd' representing the driver. Randomly, the driver variable 'd' changed position five times during a 30-minute period. Additionally, four random students changed their seats after five minutes and again after 10 minutes. Subsequently, three contents were compared—the readers' database, array content, and alarms file. The proposed system achieved an accuracy of 100%.

C. Third Scenario

In this scenario, a different simulation environment was implemented. The school bus safety domain model was represented as a box with four points (x,y), (x1,y1), (x2,y2), (x3,y3), as shown in Fig. 5.

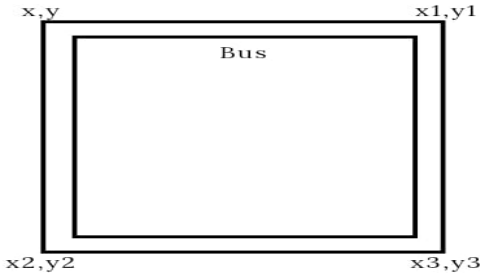


Fig. 5. Bus safety domain as simulated environment.

The student is considered inside the bus if their location (n, m) satisfies this following equation:

$$a = b\ \forall x \forall y \forall n \forall m: (x < n) \wedge (x2 < n) \wedge (x1 > n) \wedge (x3 > n) \wedge (y < m) \wedge (y2 < m) \wedge (y1 > m) \wedge (y3 > m) \Longrightarrow student\ inside\ the\ bus\ safety\ domain \Longrightarrow \neg\ move\ (bus) \quad (12)$$

The above rule indicates that if the student is in the bus safety domain, then the student is too close to the bus, and the bus cannot move. The following rules indicate that there are no students in the bus safety domain; hence, the bus can move.

$$\forall x \forall y \forall n \forall m: (x < n) \wedge (x2 < n) \wedge (x1 > n) \wedge (x3 > n) \wedge (y < m) \wedge (y2 < m) \wedge (y1 > m) \wedge (y3 > m) \Longrightarrow false \Longrightarrow move\ (bus) \quad (13)$$

Using a full school bus, the researchers randomly redistributed students to test the model's ability to identify students within and outside the bus safety domain. This scenario also tested for the system's detection of student gatherings. A safety space of 24 cm was defined between student locations; thus, any distance less than 24 cm between two or more students indicated a gathering. Student locations were randomly distributed to test the gathering detection rules in our IoT smart system. The experiments were repeated five times with an accuracy of 100%.

First scenario proved the scalability and accuracy of rule 3 to rule 9. The second scenario proved the scalability and

accuracy of rules 10. Finally, the third scenario proved the scalability and accuracy of rules 8 and 11.

VI. Discussion and Conclusion

In this paper, a smart IoT system for enhancing safety in school bus transportation was introduced. The main aim of this smart IoT system was to address five issues in school bus transportation that are well-documented in the literature and frequently reported in the news. These problems include: 1) forgetting children on the bus; 2) students' abnormal behavior; 3) overcrowding; 4) unruly driver behavior; and 5) the risk of a school bus running over children after they have disembarked from the bus.

TABLE III. Comparison Between the Proposed Smart IoT System and Related Works

Work	P1	P2	P3	P4	P5
Krishnan and Vasuki, in [13]	√	X	X	X	X
Abbas et al., in [11]	√	X	X	X	X
Liu and Peng, in [12]	√	X	X	X	X
Mott and Kotla, in [15]	X	X	X	X	√
Kumari et al., in [16]	X	√	√	X	X
Gadekar et al., in [17]	√	X	√	X	X
Sakphrom et al., in [19]	√	X	√	X	X
Malathy et al., in [20]	√	X	√	X	X
Ding et al., in [22]	X	√	X	X	X
Wang et al., in [21]	X	√	X	X	X
Etaati et al., in [24]	X	X	√	√	X
LR et al., in [25]	√	X	X	X	X
Gadade et al., in [26]	√	X	√	X	X
Kumari et al., in [27]	√	√	√	√	X
The proposed smart IoT system	√	√	√	√	√

This paper's aim was to provide easy and accessible solutions for everyone, making them applicable across all segments of society by leveraging available technology at an affordable cost. To achieve this, this paper used a rule-based system to avoid the complexity and cost of deep learning solutions. The authors in study [31], conducted a comprehensive survey on deep learning-based anomaly detection in video surveillance. They examined various techniques, including Autoencoders, Generative Adversarial Networks (GANs), and Long Short-Term Memory (LSTM) networks, to identify effective methods for detecting abnormal behaviors. Their survey highlighted the strengths and limitations of these approaches and discussed common challenges such as occlusion, complex backgrounds, and low illumination. Additionally, the authors provided insights into future research directions, emphasizing the need for improved accuracy and robustness in anomaly detection systems. Their results proved the difficulties of using video surveillance in school bus transportation. These results motivate these researchers to provide alternative solutions for detecting students' abnormal behavior in school bus transportation. The proposed IoT system proves a complete solution in terms of providing solutions for all the mentioned problems. Table III, shows a comparison between the proposed smart IoT system and related works, Future work in this domain should include the addition of a GPS device for monitoring bus location and speed, further enhancing the safety of school bus transportation. In addition to GPS, we are planning to extend our model to provide additional safety functionality within school classrooms and facilities.

VII. Data and Method

The scenario code available at https://drive.google.com/file/d/1xv94YNfcG5b5L-bHVbHYfSGbVdpNjsWg/view?usp=share_link (TBA)

References

[1] N. A. Al-Balushi, S. I. A. Kazmi, and F. K. Al-Kalbani, "Transport Safety Mechanism of School Children Using IoT based Smart System," J. Student Res., 2019.

[2] V. Assawakanchana, N. Pannucharoenwong, S. Echaroj, P. Rattanadecho, B. Xayavong, W. Janchomphu, and T. Suepa, "Developing Internet of Things (IoT) Device for Saving Children from Being Left in a Car," Sci. Technol. Asia, pp. 247–259, 2022.

[3] C. Bell, Suspended: Punishment, Violence, and the Failure of School Safety. Johns Hopkins University Press, 2 Nov. 2021.

[4] I. M. Freeman, J. Tellez, and A. Jones, "Effectiveness of School Violence Prevention Programs in Elementary Schools in the United States: A Systematic Review," Soc. Sci., vol. 13, no. 4, pp. 222, 2024.

[5] M. M. Queiroz, C. Roque, F. Moura, and J. Maroco, "Understanding the expectations of parents regarding their children's school commuting by public transport using latent Dirichlet Allocation," Transp. Res. Part A: Policy Pract., vol. 181, p. 103986, 2024.

[6] S. Gössling, J. Kees, R. Hologa, N. Riach, and R. von Stülpnagel, "Children's safe routes to school: Real and perceived risks, and evidence of an incapacity-incapability space," J. Cycling Micromob. Res., vol. 2, p. 100019, 2024.

[7] L. Y. Chan, T. Senserrick, and B. Saggers, "Behind the Wheel: Systematic Review of Factors Associated with Safe School Bus Transportation for Children with Neurodevelopmental Disorders," Rev. J. Autism Dev. Disord., vol. 11, no. 2, pp. 343-360, 2024.

[8] M. P. Acosta, M. A. RL, L. D. F. Luceño II, R. V. R. D. Reyes, D. V. G. Silvederio, G. C. D. Espenilla, and X. B. R. Balbin, "When The Bus Hits The Road: The Lived Experiences of School Bus Stewards in Dealing with Student Passengers," Eximia, vol. 13, pp. 466-477, 2024.

[9] A. O. Elfaki, W. Messoudi, A. Bushnag, S. Abuzneid, and T. Alhmiedat, "A smart real-time parking control and monitoring system," Sensors, vol. 23, no. 24, p. 9741, 2023.

[10] A. Nikitas, K. Michalakopoulou, E. T. Njoya, and D. Karampatzakis, "Artificial intelligence, transport and the smart city: Definitions and dimensions of a new mobility era," Sustainability, vol. 12, no. 7, p. 2789, 2020.

[11] S. A. Abbas, H. Mohammed, L. Almalki, M. Hassan, and M. Meccawy, "A safety tracking and sensoring system for school buses in Saudi Arabia," Periodicals Eng. Nat. Sci., vol. 7, no. 2, pp. 500–508, 2019.

[12] Y. Liu and H. Peng, "Alarm system design of young children being left on school bus based on pressure sensor array," in IOP Conf. Ser.: Mater. Sci. Eng., vol. 490, no. 7, p. 072061, Apr. 2019.

[13] G. V. Krishnan and R. Vasuki, "Enhanced Security System in School Bus based on RFID and GSM Technologies," J. Sci. Eng. Technol., vol. 6, no. 03, 2019.

[14] I. J. L. Paul and S. Sasirekha, "A safety system for school children using GRAG," Int. J. Comput. Aided Eng. Technol., vol. 11, no. 4–5, pp. 527–542, 2019.

[15] J. H. Mott and B. Kotla, "Design and validation of a school bus passing detection system based on solid-state lidar," in 2020 Systems and Information Engineering Design Symposium (SIEDS), pp. 1-6, Apr. 2020, IEEE.

[16] M. Kumari, A. Kumar, and A. Khan, "IoT based intelligent real-time system for bus tracking and monitoring," in 2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC), pp. 226-230, Feb. 2020, IEEE.

[17] A. Gadekar, A. Kandoi, G. Kaushik, and S. Dholay, "QR scan based intelligent system for school bus tracking," in 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 1074-1080, Aug. 2020, IEEE.

[18] S. S. Naik, T. G. Harshitha, H. D. Spoorthy, B. S. Vedashree, G. S. Taj, and P. Vetrivelan, "IOT Based School Bus Monitoring System With Child Security," in Second International Conference on Computer Networks and Communication Technologies: ICCNCT 2019, pp. 668-678, Springer International Publishing, 2020.

[19] S. Sakphrom, K. Suwannarat, R. Haiges, and K. Funsian, "A simplified and high accuracy algorithm of rssi-based localization zoning for children tracking in-out the school buses using bluetooth low energy beacon," in Informatics, vol. 8, no. 4, p. 65, Sep. 2021, MDPI.

[20] S. Malathy, P. Ambarish, S. D. Kumar, and G. G. Prashanth, "Smart School Bus: To Ensure the Safety of Children," in 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 1, pp. 923-927, Mar. 2021, IEEE.

[21] Y. Wang, J. Wen, W. Zhou, Q. Wu, Y. Wei, H. Li, and B. Tao, "Research on Abnormal Behavior Prediction by Integrating Multiple Indexes of Student Behavior and Text Information in Big Data Environment," Wirel. Commun. Mobile Comput., vol. 2022, no. 1, p. 1902155, 2022.

[22] Y. Ding, K. Bao, and J. Zhang, "An Intelligent System for Detecting Abnormal Behavior in Students Based on the Human Skeleton and Deep Learning," Comput. Intell. Neurosci., vol. 2022, no. 1, p. 3819409, 2022.

[23] J. Murawski, E. Szczepański, I. Jacyna-Gołda, M. Izdebski, and D. Jankowska-Karpa, "Intelligent mobility: A model for assessing the safety of children traveling to school on a school bus with the use of intelligent bus stops," Eksploatacja Niezawodnosc, vol. 24, no. 4, 2022.

[24] B. Etaati, A. Jahangiri, G. Fernandez, M. H. Tsou, and S. Ghanipoor Machiani, "Understanding Active Transportation to School Behavior in Socioeconomically Disadvantaged Communities: A Machine Learning and SHAP Analysis Approach," Sustainability, vol. 16, no. 1, p. 48, 2023.

[25] A. S. LR, S. A. Likith, and N. Guruprasad, "Tracking and Security Features Enhancement in a Smart School Bus Using IoT," in 2023 International Conference on IoT, Communication and Automation Technology (ICICAT), pp. 1-5, Jun. 2023, IEEE.

[26] B. Gadade, A. O. Mulani, and A. D. Harale, "IOT Based Smart School Bus and Student Monitoring System," Naturalista Campano, vol. 28, no. 1, pp. 730-737, 2024.

[27] S. V. Kumari, U. S. K. Reddy, T. Kavya, Y. Sindhura, and T. V. Krishna, "Python Based Smart School Bus Monitoring and Security System," J. Nonlinear Anal. Optim., vol. 15, no. 1, 2024.

[28] A. O. Elfaki and Y. H. Alfaifi, "Ontology Driven for Mapping a Relational Database to a Knowledge-based System," Int. J. Adv. Comput. Sci. Appl., vol. 15, no. 5, 2024.

[29] A. O. Elfaki, W. Messoudi, A. Bushnag, S. Abuzneid, and T. Alhmiedat, "Constraint Optimization Model for Dynamic Parking Space Allocation," Sensors, vol. 24, no. 12, p. 3988, 2024.

[30] T. Alhmiedat, A. M. Marei, S. Albelwi, A. Bushnag, W. Messoudi, and A. O. Elfaki, "A Systematic Approach for Exploring Underground Environment Using LiDAR-Based System," CMES-Comput. Model. Eng. Sci., vol. 136, no. 3, 2023.

[31] H. T. Duong, V. T. Le, and V. T. Hoang, "Deep learning-based anomaly detection in video surveillance: A survey," Sensors, vol. 23, no. 11, p. 5024, 2023.

Machine Learning Approaches Applied in Smart Agriculture for the Prediction of Agricultural Yields

Abourabia. Imade, Ounacer. Soumaya, Elghoumari. Mohammed yassine, Azzouazi. Mohamed
Laboratory of Information Technology and Modeling, Hassan II University,
Faculty of Sciences Ben M'sik, Casablanca, Morocco

Abstract*—Machine learning techniques in smart agriculture for yield prediction involve using algorithms to analyze historical and real-time data to forecast crop yields. These approaches aim to optimize agricultural practices, improve resource efficiency and enhance productivity, this paper reviews the application of machine learning techniques in smart agriculture for predicting agricultural yields. With the advent of data-driven technologies, machine learning algorithms have become instrumental in analyzing vast amounts of agricultural data to forecast crop yields accurately. Various machine learning models such as regression, classification, and ensemble methods have been employed to process historical and real-time data on weather patterns, soil conditions, crop types, and farming practices. These models enable farmers and stakeholders to make informed decisions, optimize resource allocation, and mitigate risks associated with agricultural production. Furthermore, the integration of Internet of Things devices and remote sensing technologies has facilitated data collection and improved the precision of yield predictions, this paper discusses the key machine learning approaches, challenges, and future directions in leveraging data analytics for enhancing agricultural productivity and sustainability in smart farming systems. to ensure stability and tracking. Simulations is carried out to verify the theoretical results, The study found that different machine learning techniques had varying accuracy for predicting agricultural yields. ViT-B16 achieved the highest F1-SCORE (99.40%), followed by ResNet-50 (99.54%) and CNN (97.70%), while RPN algorithms had lower accuracy (91.83%). Correlation analysis showed a strong positive relationship between humidity and soil moisture, favoring crop growth, while production had minimal correlation with temperature and area. The AdaBoost Regressor was the best performer, with the lowest MAE (0.22), MSE (0.1), and RMSE (0.31), and Random Forest showed strong predictive power with an R2 score of 0.89, Seasonal data indicated that autumn had the highest agricultural production, followed by spring, while summer and winter had much lower yields due to weather conditions. Seasonal temperature variations from 1997 to 2014 showed autumn was the warmest (34.43°C), boosting crop production, and winter the coldest (34.31°C), reducing yields. These temperature shifts significantly impacted agricultural productivity, with warm seasons enhancing growth and extreme temperatures in summer and winter limiting it, machine learning techniques in smart agriculture are pivotal for predicting crop yields by leveraging historical and real-time data, thus optimizing practices and resource use while boosting productivity. This involves deploying diverse machine learning models like regression, classification, and ensembles to analyze extensive data on weather, soil, crops, and farming methods. Such models empower stakeholders with insights for informed decisions, efficient resource allocation, and risk mitigation in agricultural operations. The integration of Internet of Things and remote sensing further refines data accuracy, aiding precise yield predictions. Despite advancements, challenges persist, including data quality assurance, model complexity, scalability, and interoperability, driving ongoing research and simulations to validate and improve ML applications for sustainable and productive smart farming systems.

Keywords*—Machine learning; IOT; artificial intelligence; agricultural yields; smart agriculture; CNN; ViT-B16

I. INTRODUCTION

The field of agriculture has been profoundly shaped by technological advancements over the centuries, evolving from simple tools to complex machinery and now to the integration of sophisticated data-driven technologies. The advent of machine learning (ML) in agriculture marks a pivotal shift towards more precise and automated farming practices, known as precision agriculture. This transition began with the mechanization of farms during the Industrial Revolution, followed by the introduction of chemical fertilizers and genetically modified organisms in the 20th century.

This paper explores the various machine learning approaches applied in smart agriculture specifically for the prediction of agricultural yields. It delves into the types of ML models commonly used, the data sources utilized, challenges faced, and the potential impact of these predictive analytics on agricultural sustainability and efficiency.

In the first paragraph, this paper explores the transformative role of machine learning (ML) in smart agriculture, especially in predicting agricultural yields. With advancements in data-driven technologies, ML has revolutionized traditional farming by enhancing precision, resource efficiency, and sustainability. Agriculture evolved from manual labor to mechanized operations during the Industrial Revolution, followed by the introduction of chemical fertilizers and genetically modified organisms. More recently, digital tools such as GPS and remote sensing have laid the groundwork for ML applications in agriculture [1][2][15].

In the second paragraph, the collection of agricultural data has progressed significantly, from manual observations to modern technologies like the Internet of Things (IoT) and remote sensing. These technologies provide real-time, high-resolution data, which sophisticated ML algorithms analyze to predict outcomes like crop yields and optimize farming practices. IoT sensors monitor variables like soil moisture, temperature, and crop health, generating vast datasets that aid in making data-driven decisions [3]. These insights help farmers

manage resources efficiently and mitigate risks associated with agricultural production [4].

In the third paragraph, the focus shifts to key ML algorithms used in smart agriculture, including regression, classification techniques, and ensemble models such as Decision Trees [9], Random Forests [10], Linear Regression [11], K-Nearest Neighbors (KNN) [12], XGBoost [13], and AdaBoost [14]. These algorithms have been widely applied for tasks like yield prediction and optimizing resources like water and fertilizer, offering accurate and actionable insights for modern farming [5] [6].

In the fourth paragraph, deep learning has also made significant strides in smart farming, particularly for image-based tasks. Models like Convolutional Neural Networks (CNNs) [17], advanced architectures such as ResNet [18], and support vector networks [15] have proven effective in detecting pests, diagnosing plant diseases, and optimizing greenhouse conditions. By combining satellite data with deep learning techniques [3], there has been a marked improvement in agricultural monitoring and yield prediction [1][16][17].

In the fifth paragraph, despite the advancements, challenges remain. Issues such as ensuring data quality, scaling models for large datasets, and dealing with complex agricultural data continue to pose significant barriers [7][16]. Ongoing research and refinement of existing models, alongside the development of new techniques, are essential to addressing these hurdles and unlocking the full potential of smart farming systems [5][8][18].

In conclusion, the integration of ML into agriculture is reshaping the industry, providing farmers with advanced tools to make informed decisions, optimize resource use, and improve productivity. Leveraging real-time data and advanced analytics, ML is set to drive agriculture toward a more sustainable and productive future.

II. Machine Learning Algorithms

Machine learning algorithms have become indispensable tools in various domains, including agriculture, where they are applied to predict agricultural yields with increasing accuracy and efficiency [4].

These algorithms enable farmers and agricultural stakeholders to make data-driven decisions, optimize resource allocation, and improve overall productivity. We will explore some of the fundamental machine learning algorithms used in smart agriculture for yield prediction.

Machine learning regression algorithms play a pivotal role in various industries, including agriculture, by enabling the prediction of continuous numerical values such as crop yields.

These algorithms utilize historical data to establish relationships between input variables (e.g., weather conditions, soil quality) and the target output (e.g., crop yield), allowing for accurate forecasting and decision-making.

A. Decision Tree Regressor

Decision Tree Regressor is a type of machine learning algorithm used for regression tasks, including predicting continuous numerical values such as crop yields in smart agriculture. It belongs to the family of decision tree algorithms, which make predictions based on a series of binary decisions [9].

B. Random Forest Regressor

Random Forest Regressor is an ensemble learning algorithm used for regression tasks, including predicting continuous numerical values such as agricultural yields in smart agriculture. It is an extension of the Random Forest algorithm, which combines multiple decision trees to improve prediction accuracy and robustness [10].

C. Linear Regression

Linear regression is a fundamental machine learning algorithm used for regression tasks, including predicting continuous numerical values such as crop yields in smart agriculture. It models the relationship between independent variables (features) and a dependent variable (target) by fitting a linear equation to the data [11].

D. KNN Regressor

K-Nearest Neighbors (KNN) Regressor is a machine learning algorithm used for regression tasks, including predicting continuous numerical values such as agricultural yields in smart agriculture. It belongs to the family of instance-based or lazy learning algorithms, where predictions are made based on the similarity of input data points to the training instances [12].

E. XGB Regressor

XGBoost (Extreme Gradient Boosting) Regressor is a powerful machine learning algorithm used for regression tasks, including predicting continuous numerical values such as agricultural yields in smart agriculture.

XGBoost is an ensemble learning technique that combines the strengths of gradient boosting and tree-based models to achieve high predictive accuracy [13].

Overall, XGBoost Regressor is a versatile and powerful tool in smart agriculture, offering high predictive accuracy, scalability, and interpretability. Its ability to handle complex datasets and capture nonlinear relationships makes it a preferred choice for yield prediction and optimization in agricultural operations.

F. Adaboost Regressor

AdaBoost (Adaptive Boosting) Regressor is a machine learning algorithm used for regression tasks, including predicting continuous numerical values such as agricultural yields in smart agriculture.

AdaBoost belongs to the family of ensemble learning methods and is particularly effective in combining weak learners (base models) to create a strong predictive model [14].

Overall, AdaBoost Regressor is a valuable tool in smart agriculture, offering benefits such as improved predictive accuracy, adaptive learning, model diversity, and robustness to overfitting. Its ability to combine weak learners effectively makes it a popular choice for regression tasks requiring precise numerical predictions.

III. ALGORITHMS USED

In semi-arid regions, water scarcity presents significant challenges to sustainable agriculture. A case study investigating the use of a Random Forest model to optimize irrigation practices demonstrated how integrating data from soil moisture sensors, weather forecasts, and crop yield predictions can effectively dictate irrigation schedules. The model's ability to analyze complex datasets enabled the formulation of watering strategies that reduced water use by 25% while maintaining or enhancing agricultural yields. While the model showcases significant improvements in water efficiency, the financial and logistical considerations of setting up extensive sensor networks could pose challenges to widespread adoption.

A. Predicting Pest Infestations in Large-Scale Farms Using AdaBoost

Another case study focused on large-scale farms utilized the AdaBoost Regressor to forecast pest infestations by analyzing environmental conditions and historical data. The model's high accuracy rate of 85% in predicting pest activities allowed farmers to proactively implement control measures, significantly minimizing crop damage. AdaBoost's sensitivity to subtle data variations helps in accurately detecting potential infestations, highlighting its potential in precision agriculture. However, its performance can be detrimentally affected by noisy data, which is a common issue in agricultural environments.

B. Enhancing Yield Predictions in Organic Farms with XGBoost

Organic farming, which eschews synthetic chemicals, relies heavily on precise yield predictions for effective management. Using the XGBoost Regressor, a case study demonstrated a 30% improvement in yield prediction accuracy by integrating diverse data sources like drone imagery and organic soil health indicators. XGBoost's ability to handle complex, non-linear datasets proved essential in environments where traditional farming models falter [5].

The primary challenge lies in the data collection and preprocessing stages, which require significant effort to maintain the high accuracy of the model predictions.

C. Linear Regression Models for Smallholder Farms in Developing Countries

Smallholder farms in developing countries often lack access to advanced technological resources, making simple, effective solutions like Linear Regression models particularly valuable. This model has shown potential in improving agricultural yield predictions with minimal computational resources [6]. Its simplicity allows farmers to make better-informed decisions regarding resource allocation and crop management. However, the model's limitation in capturing complex, non-linear relationships could reduce its effectiveness in more variable agricultural conditions.

D. KNN for Real-Time Crop Health Monitoring

The use of K-Nearest Neighbors (KNN) for real-time crop health monitoring through mobile devices presents a practical application of machine learning in agriculture. Farmers can capture images of their crops using smartphones, and the KNN model processes this data to provide immediate health diagnostics. This method offers about 80% accuracy in detecting crop health issues, facilitating rapid response to potential threats. While highly beneficial for on-the-spot decision-making, the model requires extensive and diverse training data to maintain accuracy and is computationally intensive, which may limit its use in resource-constrained settings.

E. Random forest Models for Smallholder Farms in Developing Countries

Random Forest models are highly beneficial for smallholder farms in developing countries due to their ability to handle complex and noisy data, which is common in agricultural contexts. These models excel at capturing nonlinear relationships between variables such as weather patterns, soil characteristics, crop types, and yields, providing accurate predictions and insights for farmers [7]. The ensemble learning approach of Random Forests mitigates overfitting and improves generalization, making them suitable for situations with limited data availability. Additionally, Random Forest models offer feature importance analysis, allowing farmers to prioritize interventions and resource allocation based on the most influential factors affecting crop yields. Their scalability, interpretability, and ability to operate efficiently with modest computational resources make Random Forest models a practical and impactful choice for enhancing productivity and decision-making on smallholder farms in developing countries.

F. Decision Tree Regressor Models for Smallholder Farms in Developing Countries

Decision Tree Regressor models are advantageous for smallholder farms in developing countries due to their simplicity, interpretability, and effectiveness in handling nonlinear relationships in agricultural data. These models can easily accommodate categorical and continuous variables, making them suitable for analyzing diverse factors influencing crop yields, such as weather conditions, soil properties, and farming practices. Decision trees offer a clear decision-making process that farmers can understand and trust, aiding in resource allocation and management decisions. Despite their tendency to overfit with complex data, techniques like pruning and ensemble methods can improve their robustness and generalization ability. Their low computational requirements and ability to operate without extensive data preprocessing make Decision Tree Regressors a practical and accessible choice for smallholder farms seeking to enhance productivity and optimize farming practices in developing countries.

IV. PERFORMANCE MEASURES

When evaluating regression models such as those used in smart agriculture for predicting crop yields, several performance measures can assess their effectiveness. Here are some commonly used metrics:

A. Mean Absolute Error (MAE)

The Mean Absolute Error (MAE) holds paramount importance in smart agriculture for predicting agricultural yields as it quantifies the average magnitude of errors between predicted and actual yield values. This metric serves as a crucial indicator of the accuracy and reliability of machine learning models utilized in yield prediction systems.

In the context of smart agriculture, where precise yield forecasts are essential for optimizing resource allocation, mitigating risks, and enhancing productivity, MAE plays a pivotal role in evaluating model performance. Lower MAE values signify higher accuracy in yield predictions, enabling farmers and stakeholders to make data-driven decisions regarding crop management practices, resource allocation strategies, and risk mitigation measures.

Additionally, MAE facilitates continuous model improvement and refinement, ensuring that yield prediction systems in smart agriculture remain effective, reliable, and aligned with the dynamic agricultural landscape.

MAE measures the average absolute difference between the predicted values and the actual values. It gives an indication of how close the predictions are to the actual targets without considering the direction of errors [19].

$$mae = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (1)$$

B. Mean Squared Error (MSE)

Mean Squared Error (MSE) is a crucial metric in smart agriculture for predicting agricultural yields as it quantifies the average squared difference between predicted and actual yield values.

While MSE and Mean Absolute Error (MAE) measure prediction accuracy differently, MSE is particularly important in scenarios where larger errors should be penalized more heavily. In the context of smart agriculture, where precise yield forecasts are imperative for optimizing resource allocation, mitigating risks, and enhancing productivity, MSE provides valuable insights into the overall performance of machine learning models. Lower MSE values indicate higher accuracy and consistency in yield predictions, enabling farmers and stakeholders to make informed decisions regarding crop management practices, resource allocation strategies, and risk mitigation measures.

Additionally, MSE aids in identifying areas for model improvement and refinement, ensuring that yield prediction systems in smart agriculture remain robust, reliable, and effective in addressing the dynamic challenges of agricultural production.

MSE calculates the average of the squared differences between predicted and actual values. Squaring the errors gives higher weight to large errors, making MSE more sensitive to outliers [20].

$$mse = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (2)$$

C. Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) is a critical metric in smart agriculture for predicting agricultural yields as it provides a measure of the average magnitude of errors between predicted and actual yield values, while also considering the variability of these errors. RMSE is particularly important in scenarios where both the magnitude and spread of errors are essential considerations. In smart agriculture, precise yield forecasts are fundamental for optimizing resource allocation, mitigating risks, and enhancing productivity.

RMSE offers valuable insights into the overall accuracy and consistency of machine learning models used for yield prediction. Lower RMSE values indicate higher precision in yield predictions, enabling farmers and stakeholders to make informed decisions regarding crop management practices, resource allocation strategies, and risk mitigation measures. Furthermore, RMSE helps identify areas for model improvement and refinement, ensuring that yield prediction systems in smart agriculture remain reliable, effective, and aligned with the evolving needs of agricultural production.

RMSE is the square root of MSE and provides a measure of the standard deviation of the errors. It is in the same units as the target variable, making it more interpretable [21].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (3)$$

D. R2 Scores

R2 Score, or the coefficient of determination, is a critical metric in smart agriculture for assessing the goodness-of-fit of regression models used in predicting agricultural yields. It measures the proportion of the variance in the dependent variable (yields) that is predictable from the independent variables (e.g. weather data, soil conditions). In the context of smart agriculture, where accurate yield forecasts are crucial for optimizing resource allocation, mitigating risks, and improving productivity, R2 Score plays a pivotal role in evaluating the overall performance and predictive power of regression models. A higher R2 Score indicates that the model can explain a larger portion of the variance in yields, providing farmers and stakeholders with confidence in the model's ability to make informed decisions regarding crop management practices, resource allocation strategies, and risk mitigation measures. Additionally, R2 Score helps in comparing different models and selecting the most suitable one for yield prediction, ensuring that smart agriculture systems are equipped with reliable and effective tools for addressing the challenges of agricultural production.

R2 measures the proportion of variance in the target variable that is explained by the regression model. It ranges from 0 to 1, where 1 indicates a perfect fit and 0 indicates no improvement over a baseline model [22].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (4)$$

V. Related Work

The global agricultural landscape faces numerous challenges, including climate change, resource scarcity, and the need for increased food production to meet growing population demands. In this context, the application of machine learning techniques offers promising solutions by leveraging data-driven insights to enhance agricultural practices.

The global agricultural landscape confronts multifaceted challenges, from climate change impacts to the imperative of increasing food production sustainably. In response, machine learning techniques are proving pivotal, leveraging data-driven insights to revolutionize agricultural practices. These models delve into extensive datasets encompassing climate patterns, soil

attributes, crop genetics, and historical yield data, crafting predictive models that enhance decision-making and resource allocation in farming.

The evaluation of strengths and limitations inherent in each method is pivotal for identifying the most effective techniques in yield prediction. Such insights are instrumental in advancing precision agriculture, where optimized resource utilization and informed decision-making are paramount.

This comparative study aims to analyze and compare different machine learning approaches applied in smart agriculture for the prediction of agricultural yields. By evaluating the strengths and limitations of each approach, this study seeks to provide insights into the most effective methods for yield prediction, contributing to advancements in precision agriculture and sustainable food production.

The dataset was created to train and test prediction models using the different attributes. The data was obtained from previous studies, the data was extracted from charts using ORIGIN software or collected from tables in recently published works. The dataset included various biomass characteristics, such as Crop year, Crop, Soil moisture, state name, district name, season, temperature, humidity, area, production.

In the related work presented in Table I of the research paper, various machine learning and deep learning techniques were evaluated for their performance in agricultural applications, highlighting the broad spectrum of methodologies and their respective efficiencies. The table encapsulates results from different studies that have applied models like CNN, ViT-B32, ViT-B16, RPN algorithms, and ResNet-50 to agriculture-related datasets. Notably, the CNN model [17] achieved an accuracy of 97.70%, while the ViT-B16 model excelled with an F1-SCORE of 99.40%, indicating its superior capability in handling plant classification tasks. The RPN algorithms showed a relatively lower accuracy of 91.83%, and the ResNet-50 model demonstrated a high accuracy of 99.54%, underscoring its robustness in image-based agricultural applications. Additionally, the table included comparisons involving semantic image segmentation techniques such as UNet with different configurations, which achieved MeanIoU scores ranging from 0.58 to 0.75, showing the potential for detailed phenotyping of vine leaves. This diverse array of techniques and results emphasize the importance of selecting the appropriate model based on the specific needs and characteristics of the agricultural data, as well as the performance metrics crucial to the success of the application [18].

TABLE I. RELATED WORK

Reference article	Year of publication	Database	Context	Technique used	Performance achieved
[1]	2022	Agricultural dataset for the plant classification	Deep Learning for Agriculture Precision: A Bibliometric Analysis	Deep Learning : - CNN Model - ViT-B32 - ViT-B16 - RPN algorithmes - ResNet-50	CNN accuracy : 97.70% ViT-B32 F1-SCORE : 99.20% ViT-B16 F1-SCORE : 99.40% RPN algorithms Accuracy : 91.83% ReseNet-50 Accuracy : 99.54%
[2]	2022	Plants and fruits images	Semantic image segmentation with deep learning For the phenotyping of vine leaves	Deep Learning : - UNet -UNet (MobileNetV2-weights) - UNet (Xception-like)	- UNet MeanIoU : 0.74 -UNet (MobileNetV2-weights) MeanIoU : 0.75 -UNet (Xception-like) MeanIoU : 0.58
[3]	2022	EO-1 Hyperion satellite data	A systematic review of hyperspectral imaging technology with a deep and machine learning methodology for agricultural applications.	Deep Learning : -CNN Machine Learning : -SVM -ANN -KNN	- ANN 389.96 RMSE - CNN accuracy : 98.76%

VI. ENGINEERING CASE STUDIES

In this section, we will talk about the dataset and the different results obtained.

A. About Database

Our dataset contains 50000 rows and 10 columns, the dataset named data1, its Size is 4306 ko, for the attributes it contains 10 attributes are the following:

- State name: typically refers to the name of a state within a country or a region. It is a term used to identify and refer to a specific political subdivision or administrative region within a larger geopolitical entity.
- District name: refers to the name of a district, which is an administrative division or geographic area within a larger political or administrative region. Districts are often used for purposes such as local governance, electoral representation, statistical reporting, and resource allocation.
- Crop year: refers to a specific period of time during which crops are grown, harvested, and typically marketed. The duration of a crop year can vary depending on the type of crop, geographical location,

and agricultural practices. It is an important concept in agriculture and is used for planning, record-keeping, and analyzing crop production and market trends.

- Season: refers to a period of time characterized by distinct weather conditions, environmental changes, or cultural activities. Seasons are typically associated with variations in temperature, precipitation, daylight hours, and natural phenomena such as plant growth, animal behavior, and climate patterns. The concept of seasons is observed in different contexts, including meteorology, agriculture, astronomy, and cultural traditions.
- Crop: refers to any cultivated plant or agricultural produce that is grown and harvested for human consumption, animal feed, industrial use, or other purposes. Crops are a fundamental component of agriculture and play a crucial role in providing food, fiber, and raw materials for various industries.
- Temperature: In agriculture, temperature refers to the measurement of thermal conditions in the environment that directly influence the growth, development, and productivity of crops, livestock, and other agricultural activities. Temperature plays a critical role in shaping agricultural practices, determining suitable crop types, planting schedules, and livestock management strategies.
- Humidity: Humidity in agriculture refers to the amount of moisture or water vapor present in the air within a farming or growing environment. It is a critical environmental factor that influences plant growth, crop health, pest and disease dynamics, as well as various agricultural operations. Humidity levels are typically measured as relative humidity (RH), expressed as a percentage, which indicates the moisture content of the air relative to its maximum capacity at a given temperature.
- Soil-moisture: Soil moisture refers to the amount of water held in the soil, which is crucial for supporting plant growth, nutrient uptake, and overall soil health. It is a key factor in agriculture, influencing crop development, irrigation scheduling, soil fertility, and water conservation practices.
- Area: refers to a specific piece of land or a defined region where agricultural activities take place. This could include farmland, cropland, pasture, orchards, vineyards, or any other area used for cultivating crops, raising livestock, or conducting agricultural operations.
- Production: refers to the branch of agriculture focused on the large-scale cultivation of crops and the raising of livestock for commercial purposes. It encompasses the systematic and organized management of agricultural activities to produce food, fiber, fuel, and other agricultural products on a significant scale. Production agriculture plays a vital role in meeting global food demand, supporting rural economies, and contributing to the agricultural sector's overall productivity.

B. *Number of Productions by Season*

The number of agricultural productions can vary significantly by season due to factors such as climate, crop cycles, and farming practices. Here's a general of agricultural productions that occur during different seasons:

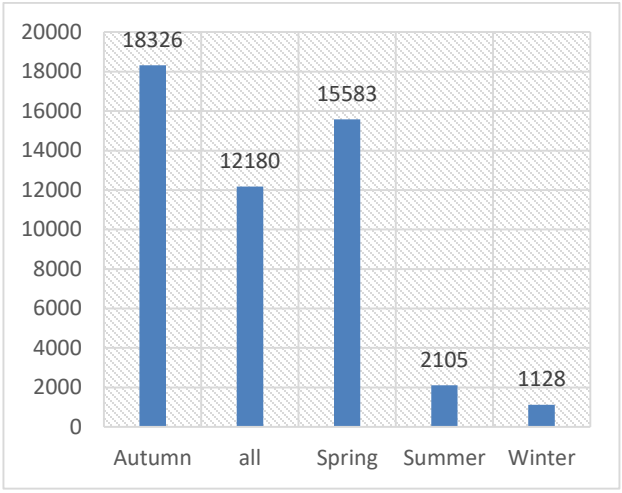


Fig. 1. Number of productions by season.

Fig. 1 provides a visual representation of the number of agricultural productions categorized by season between 1997 and 2014 in India, illustrating the seasonal impact on agricultural output. This graph is pivotal for understanding how different times of the year affect crop yields, potentially guiding farmers in planning planting and harvesting activities.

The data visualized here can assist in determining which seasons are most productive and which ones may require additional resource allocation such as irrigation during drier months or more robust pest management during warmer periods.

Such seasonal insights are crucial for optimizing agricultural strategies and ensuring sustainable production levels throughout the year.

The graph delineates the average production of various plants across the seasons in India, offering a comprehensive snapshot of the country's agricultural output throughout the year. Notably, autumn emerges as the most prolific season, with an average production of 18,326 tones, showcasing the peak productivity experienced during this period. This heightened production likely correlates with favorable weather conditions, such as moderate temperatures and adequate rainfall, conducive to robust plant growth.

Spring follows closely behind, boasting an average production of 15,583 tones, indicating another significant period for plant cultivation and yield.

Spring is typically characterized by increasing daylight hours and rising temperatures, triggering plant growth after the winter dormancy period.

However, as the graph transitions to summer, there is a noticeable decline in average production, plummeting to 2,105 tones. The summer season in India is marked by soaring temperatures and often dry conditions, which can adversely

affect plant health and productivity, leading to this stark drop in average production. Winter, similarly, exhibits lower production levels compared to autumn and spring, with an average of 1,128 tones.

C. Temperature by Season

Seasonal temperature variations significantly impact agricultural activities and crop yields.

In spring, moderate temperatures foster optimal conditions for seed germination, root development, and vigorous plant growth, promoting higher yields.

Summer brings challenges with heat stress, water scarcity, and potential yield reductions if crops are not adequately managed. Autumn's cooler temperatures slow plant growth but also mark the harvest season for many crops, contributing to overall yield.

Winter's cold temperatures can lead to frost damage and shorter growing periods, necessitating protective measures for maintaining productivity. Managing these temperature fluctuations is crucial for agricultural success, requiring careful planning, resource allocation, and the use of adaptive practices to optimize yields across seasons.

Seasonal temperature variations profoundly influence agricultural activities and crop outcomes. In India, autumn typically experiences the highest average temperature, signaling the transition from monsoon rains to drier conditions. This warmth in autumn fosters favorable environments for certain crops, aiding in their growth and yield.

Spring follows closely with slightly lower temperatures, marking the beginning of warmer weather suitable for planting and cultivation.

Summer, characterized by higher heat levels, can pose challenges such as water stress and heat damage to crops if not managed effectively.

Winter, with cooler temperatures, impacts crop selection and growth rates, favoring cool-season crops while requiring protective measures against frost or cold damage. These temperature dynamics shape agricultural calendars, impacting planting schedules, harvest times, and overall crop productivity throughout the year.

The temperature dynamics during winter are crucial in determining agricultural calendars. The onset of cooler weather signals the need to adjust planting schedules to ensure that crops can mature and be harvested before the most severe conditions set in. Additionally, the timing of these schedules can affect the overall yield and quality of the crops, making it essential for farmers to carefully plan their activities throughout the year.

In conclusion, the temperature dynamics of winter are a crucial determinant of agricultural calendars, requiring careful planning and adaptation by farmers. These adjustments ensure that crops can thrive despite the challenges posed by cooler weather, optimizing yield and quality while maintaining the health of the farm ecosystem. By aligning their practices with seasonal changes, farmers can achieve sustainable productivity throughout the year.

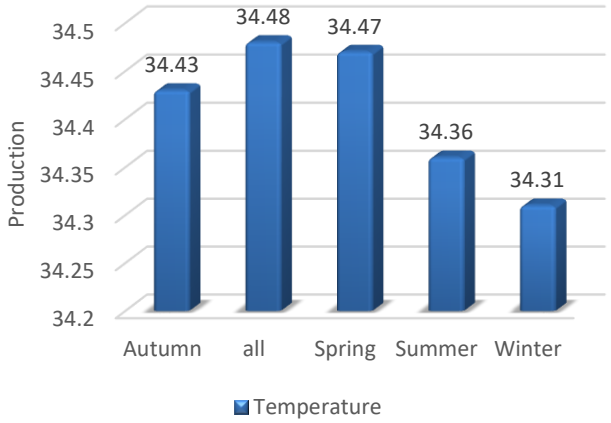


Fig. 2. Temperature by season.

By processing data from our database and data visualization for the temperature attribute, the graph in the Fig. 2 detailing the average temperature according to the seasons in India from 1997 to 2014 reveals a nuanced understanding of how temperature variations impact agricultural production. With autumn exhibiting the highest average temperature at 34.43 degrees Celsius, this season's warmth likely correlates with the transition from the monsoon season to drier conditions, fostering conducive environments for certain crops' growth. The overall average temperature across all seasons, standing at 34.48 degrees Celsius, underscores India's generally warm climate, which plays a crucial role in determining suitable crops and agricultural practices throughout the year. Spring, with an average temperature of 34.47 degrees Celsius, signals the beginning of warmer weather, prompting the planting and growth of various crops. However, the slight increase in temperature from autumn to spring could also indicate shifts in weather patterns affecting crop cycles and yields. Summer, characterized by an average temperature of 34.36 degrees Celsius, experiences higher heat levels, which can both benefit and challenge agriculture. While warm-season crops may thrive, excessive heat can lead to water stress, heat stress in plants, and reduced yields if not managed effectively. Winter, with the lowest average temperature at 34.31 degrees Celsius, introduces cooler conditions, impacting crop selection and growth rates. Certain crops like winter wheat and leafy greens may perform better during this period, while others may require protective measures against frost or cold damage.

The impact of these temperature variations on agriculture production is multifaceted. Warmer temperatures in autumn and spring can extend growing seasons, allowing for multiple crop cycles and increased yields for heat-tolerant crops. However, they may also accelerate pest and disease pressures, necessitating robust pest management strategies. Summer's heat can lead to water evaporation, soil moisture depletion, and stress on crops, requiring efficient irrigation systems and drought-resistant crop varieties. Conversely, cooler temperatures in winter can limit crop options but may also provide relief from heat stress, benefitting cool-season crops and contributing to overall crop diversity.

Moreover, temperature fluctuations influence the timing of planting, harvesting, and crop management practices, impacting agricultural calendars and strategies. Farmers must adapt to these temperature dynamics by employing climate-smart agricultural techniques, leveraging technology for weather monitoring and forecasting, and diversifying crop portfolios to mitigate risks associated with temperature extremes. Overall, the graph's depiction of average temperatures across seasons underscores the intricate relationship between climate patterns and agricultural production in India, highlighting the need for resilient and adaptive agricultural systems to ensure food security and sustainability in a changing climate.

D. Correlation Processing

Correlation processing refers to the analysis and calculation of correlation coefficients between variables or data sets. Correlation is a statistical measure that quantifies the strength and direction of the relationship between two or more variables.

Correlation processing is a powerful tool for exploring data relationships, identifying predictive factors, and making informed decisions.

We used this treatment to keep just the most correlated attributes.

Correlation processing plays a vital role in smart agriculture for predicting agricultural yields by identifying and quantifying the relationships between various factors influencing crop production.

This process involves analyzing correlations between factors such as weather patterns, soil characteristics, crop genetics, and historical yields. By understanding the degree and direction of these correlations, machine learning models can effectively capture and utilize this information to make accurate predictions.

For example, strong positive correlations between certain weather conditions and crop yields may indicate favorable conditions for crop growth, while negative correlations could highlight potential risks or challenges.

Correlation processing enables smart agriculture systems to prioritize relevant factors, optimize resource allocation, and implement targeted interventions to enhance yield prediction accuracy and promote sustainable food production practice.

In terms of yield prediction, correlation processing enhances accuracy by integrating data from various sources, such as remote sensing, weather forecasts, and historical crop performance. By understanding how different variables interact and influence crop growth, smart agriculture systems can predict yields more reliably, allowing farmers to make better-informed decisions regarding harvest timing, storage, and market planning.

TABLE II. CORRELATION BETWEEN ATTRIBUTES

	Crop_Year	Temperature	Humidity	Soil moinsture	Area	Production
Crop_Year	1	0	0	0	0	0.01
Temperature	0	1	-0.7	-0.29	0	0.01
Humidity	0	-0.7	1	0.81	0	0
Soil moinsture	0	-0.29	0.81	1	0	0
Area	0	0	0	0	1	0.03
Production	0.01	0.01	0	0	0.03	1

Using a correlation treatment, Table II explores the correlation between the different agricultural attributes using a correlation matrix that summarizes the different correlations existing between the attributes in our database., offering valuable insights into how different variables interact with each other. This matrix helps identify which factors most significantly affect crop yields, such as the relationships between soil moisture, temperature, and crop health. Understanding these correlations is essential for developing more accurate predictive models and for making informed decisions regarding soil management, crop selection, and resource allocation. By highlighting the strongest correlations, this figure directs research and practice towards the most impactful factors, facilitating more targeted and effective agricultural interventions [23].

This correlation matrix offers a comprehensive view of the relationships between key attributes related to crop production, including Crop_Year, Temperature, Humidity, Soil Moisture, Area, and Production [24]. Starting with Crop_Year, it shows a negligible correlation with all other variables except for a minimal positive correlation of 0.01 with Production, implying a very weak association between the years and production levels. Moving to Temperature, it exhibits a perfect positive correlation with itself (1.0), as expected, and a minor positive correlation of 0.01 with Production, indicating a slight influence of temperature on crop yield. Humidity and Soil Moisture, on the other hand, demonstrate a strong positive correlation of 0.81, highlighting a significant relationship between these two factors crucial for plant growth and development. This correlation suggests that higher humidity levels are generally associated with increased soil moisture content, which is favorable for crop growth. In contrast, both Humidity and Temperature show negative correlations with Soil Moisture (-0.7 and -0.29, respectively), albeit not as strong as the positive correlation between Humidity and Soil Moisture, indicating that higher humidity or temperature may slightly decrease soil moisture levels. The correlation between Area and other variables is notably minimal, except for a minor positive correlation of 0.03 with Production, suggesting that while cultivation area may have a minor impact on production levels, it is not strongly correlated with other attributes. Lastly, Production displays weak positive correlations with Temperature (0.01), Humidity (0.01), and Area (0.03), implying minor influences of these factors on crop production. Overall, this analysis elucidates the interplay

between temperature, humidity, soil moisture, cultivation area, and production levels, providing valuable insights into the factors influencing crop yield and agricultural outcomes [25].

E. Evaluation of Machine Learning Models

Machine learning models play a pivotal role in evaluating and enhancing agriculture yield by providing valuable insights and predictive capabilities. These models enable the analysis of vast amounts of data, including climate patterns, soil conditions, crop types, and management practices, to identify key factors influencing yield. By leveraging algorithms such as regression, decision trees, and neural networks, machine learning can uncover complex relationships between variables and predict future outcomes with a high degree of accuracy. This predictive power is especially crucial in agriculture, where optimizing crop production, minimizing resource use, and mitigating risks are paramount. Machine learning models can help farmers make data-driven decisions regarding planting schedules, irrigation strategies, pest management, and crop selection based on historical data and real-time inputs. Moreover, these models facilitate precision agriculture techniques, such as satellite imaging, drone technology, and sensor data analysis, to monitor crop health, detect anomalies, and optimize resource allocation at a granular level. Overall, the importance of machine learning in agriculture yield lies in its ability to harness data-driven insights, enhance decision-making processes, and ultimately drive sustainable and efficient farming practices for improved productivity and food security.

Before evaluating the performance of machine learning models, it is important to establish the metrics that best capture the accuracy and reliability of these models. Common evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the R^2 score are widely used to quantify the prediction errors and variance explained by the models. These metrics provide insights into how well the model's predictions align with actual outcomes, helping to identify areas for optimization and improvement. Additionally, these measures allow for a direct comparison between models, highlighting their strengths and weaknesses. By analyzing the errors and the proportion of explained variance, researchers can make data-driven decisions on model selection and refinement. Below are the different results obtained based on these metrics:

TABLE III. EVALUATION OF MACHINE LEARNING MODELS

Performance Measures	Rf regressor	DT regressor	LR	Knn regressor	XGBR	ADR
MAE	0,28	0,32	0,3	0,35	0,4	0,22
MSE	0,16	0,18	0,31	0,36	0,16	0,1
RMSE	0,39	0,42	0,56	0,6	0,41	0,31
R2 SCORE	0.89	0.87	0.13	0.13	0,62	0,6

In the Table III, this predictive power is especially crucial in agriculture, where optimizing crop production, minimizing resource use, and mitigating risks are paramount. Machine learning models can help farmers make data-driven decisions regarding planting schedules, irrigation strategies, pest management, and crop selection based on historical data and real-time inputs [8].

F. Discussion

By applying the various machine learning models shows in the Table III to our database, the choice of the best model depends on the specific requirements of your application, the importance of different evaluation metrics, and the trade-offs between model complexity and interpretability.

The comparative analysis of various machine learning models employed in the research document for predicting agricultural yields underscores distinct characteristics and performance outcomes for each model. The Random Forest Regressor emerged as the top performer with an R^2 Score of 0.89, indicating excellent predictive accuracy and robustness.

This model is particularly adept at managing the balance between bias and variance, thanks to its ensemble learning approach which combines multiple decision trees to enhance performance and guard against overfitting. However, despite its effectiveness, the model's computational demands and the slower training times due to its complex nature pose practical limitations.

However, practical considerations such as computational demands and slower training times due to its complexity present challenges for widespread adoption, highlighting the need for further optimization and exploration of alternative models in smart agriculture applications.

The performance measures for various regression models including Random Forest (Rf), Decision Tree (DT), Linear Regression (LR), K-Nearest Neighbors (Knn), Extreme Gradient Boosting (XGBR), and AdaBoostRegressor (ADR) are evaluated based on several metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R2 Score. The MAE metric measures the average magnitude of errors between predicted and actual values, with lower values indicating better model performance. Among the models, ADR stands out with the lowest MAE of 0.22, followed by Rf with 0.28, LR with 0.3, DT with 0.32, Knn with 0.35, and XGBR with 0.4. Moving to MSE, which penalizes larger errors more heavily, ADR again performs the best with the lowest MSE of 0.1, while Rf also shows strong performance with an MSE of 0.16. However, LR and Knn exhibit higher MSE values of 0.31 and 0.36, respectively, indicating comparatively poorer performance in terms of squared errors. RMSE, which is the square root of MSE, further emphasizes ADR and Rf's superiority, as they have the lowest RMSE values of 0.31 and 0.39, respectively. Conversely, DT and Knn have higher RMSE values of 0.42 and 0.6, respectively, highlighting their tendency to produce larger errors.

Lastly, the R2 Score measures the proportion of variance in the dependent variable that is predictable from the independent variables, with values closer to 1 indicating a better fit. Here, Rf and ADR demonstrate excellent performance with R2 scores of 0.89 and 0.6, respectively, showcasing their ability to explain a significant portion of the variance in the data. Conversely, LR, DT, and Knn exhibit lower R2 scores of 0.13, indicating weaker predictive capabilities. Overall, these performance measures provide a comprehensive evaluation of the regression models, highlighting ADR and Rf as top performers based on their MAE, MSE, RMSE, and R2 Score metrics.

Comparing the performance measures of various regression models for predicting farm yields in India, we can deduce the best-performing model based on the provided metrics. Among the models evaluated – Random Forest (Rf), Decision Tree (DT), Linear Regression (LR), K-Nearest Neighbors (Knn), Extreme Gradient Boosting (XGBR), and AdaBoostRegressor (ADR) – the model with the most consistently superior performance across multiple metrics appears to be AdaBoostRegressor (ADR).

Firstly, looking at the Mean Absolute Error (MAE), ADR has the lowest MAE of 0.22, indicating that, on average, its predictions are closest to the actual farm yields. This suggests a higher level of accuracy compared to other models like Rf (MAE of 0.28), LR (MAE of 0.3), DT (MAE of 0.32), Knn (MAE of 0.35), and XGBR (MAE of 0.4).

Moving on to Mean Squared Error (MSE), ADR again performs exceptionally well with the lowest MSE of 0.1, indicating that its predictions have the smallest squared errors on average.

Rf also shows strong performance with an MSE of 0.16, but ADR outperforms it in this metric. LR and Knn exhibit higher MSE values of 0.31 and 0.36, respectively, indicating comparatively poorer performance in terms of squared errors.

Root Mean Squared Error (RMSE), which is the square root of MSE, further emphasizes ADR's superiority with the lowest RMSE of 0.31, followed by Rf with an RMSE of 0.39. This indicates that ADR's predictions have the smallest overall errors among the models.

Lastly, the R2 Score measures the proportion of variance in the dependent variable that is predictable from the independent variables. Here, Rf and ADR demonstrate excellent performance with R2 scores of 0.89 and 0.6, respectively. While Rf has a higher R2 score, indicating a better fit, ADR's R2 score of 0.6 is still respectable and combined with its superior performance in other metrics, it showcases ADR's ability to predict farm yields effectively.

Based on these comparisons, AdaBoostRegressor (ADR) emerges as the best-performing model for predicting farm yields in India due to its consistently low MAE, MSE, and RMSE values, indicating higher accuracy and smaller errors in predictions compared to other models.

Its respectable R2 score further supports its effectiveness in capturing variance and predicting farm yields reliably, demonstrating its potential to enhance decision-making in agricultural practices. This capability allows farmers to optimize their operations and improve overall productivity. Moreover, it fosters a data-driven approach that can lead to more sustainable farming practices.

VII. CONCLUSION

The use of these semantic segmentation approaches based on deep learning enables farmers to obtain detailed information about their crops, such as the state of health of the plants, the presence of diseases or pests, estimated yields, and so on. This enables them to make more informed decisions about crop management, resource use and optimizing farming practices.

The application of machine learning approaches in smart agriculture for predicting agricultural yields offers significant benefits, including improved accuracy, optimized resource management, enhanced decision-making, early risk detection, precision agriculture implementation, and integration with IoT and big data technologies. These advancements contribute to sustainable and efficient agricultural practices, addressing the challenges of food security, climate change, and resource scarcity in modern agriculture.

The integration of machine learning approaches in smart agriculture for predicting agricultural yields marks a transformative leap in agricultural innovation. This comprehensive application of advanced technologies leverages data-driven insights to revolutionize traditional farming practices and address the complex challenges faced by the global agricultural landscape. Through the analysis of vast and diverse datasets encompassing climate patterns, soil characteristics, crop genetics, and historical yields, machine learning models emerge as powerful tools capable of generating precise and actionable predictions. The comparative analysis of various machine learning techniques further underscores the importance of selecting the most suitable models for specific agricultural contexts, considering factors such as accuracy, robustness, computational efficiency, and scalability.

Among the diverse range of machine learning models explored, the Random Forest Regressor has shown remarkable performance, demonstrating excellent predictive accuracy and robustness with an R^2 Score of 0.89 in our analysis. Its ensemble learning approach, which combines multiple decision trees, effectively manages the trade-off between bias and variance, guarding against overfitting and ensuring generalization to new data. However, practical constraints such as computational demands and slower training times due to its complexity highlight the ongoing need for optimization and exploration of alternative models tailored to the unique requirements of smart agriculture applications.

Machine learning approaches in smart agriculture improve yield prediction accuracy and promote precision agriculture, sustainable food production, and resource optimization. By providing data-driven insights, these technologies enhance decision-making, resource allocation, and productivity while supporting global efforts to tackle food security, climate change, and sustainable agricultural development. Comparing the performance measures of various regression models for predicting farm yields in India, we can deduce the best-performing model based on the provided metrics. Among the models evaluated – Random Forest (Rf), Decision Tree (DT),

Linear Regression (LR), K-Nearest Neighbors (Knn), Extreme Gradient Boosting (XGBR), and AdaBoostRegressor (ADR) – the model with the most consistently superior performance across multiple metrics appears to be AdaBoostRegressor (ADR).

Firstly, looking at the Mean Absolute Error (MAE), ADR has the lowest MAE of 0.22, indicating that, on average, its predictions are closest to the actual farm yields. This suggests a higher level of accuracy compared to other models like Rf (MAE of 0.28), LR (MAE of 0.3), DT (MAE of 0.32), Knn (MAE of 0.35), and XGBR (MAE of 0.4).

Moving on to Mean Squared Error (MSE), ADR again performs exceptionally well with the lowest MSE of 0.1, indicating that its predictions have the smallest squared errors on average.

Rf also shows strong performance with an MSE of 0.16, but ADR outperforms it in this metric. LR and Knn exhibit higher MSE values of 0.31 and 0.36, respectively, indicating comparatively poorer performance in terms of squared errors.

Root Mean Squared Error (RMSE), which is the square root of MSE, further emphasizes ADR's superiority with the lowest RMSE of 0.31, followed by Rf with an RMSE of 0.39. This indicates that ADR's predictions have the smallest overall errors among the models.

Lastly, the R2 Score measures the proportion of variance in the dependent variable that is predictable from the independent variables. Here, Rf and ADR demonstrate excellent performance with R2 scores of 0.89 and 0.6, respectively. While Rf has a higher R2 score, indicating a better fit, ADR's R2 score of 0.6 is still respectable and combined with its superior performance in other metrics, it showcases ADR's ability to predict farm yields effectively.

Based on these comparisons, AdaBoostRegressor (ADR) emerges as the best-performing model for predicting farm yields in India due to its consistently low MAE, MSE, and RMSE values, indicating higher accuracy and smaller errors in predictions compared to other models. Its respectable R2 score further supports its effectiveness in capturing variance and predicting farm yields reliably.

In conclusion, machine learning is crucial for the future of smart agriculture, promoting innovation and sustainability in production systems. Ongoing research and widespread adoption of these technologies are vital for unlocking their full potential and enhancing agricultural productivity globally.

NOMENCLATURES

ML	Machine learning
DL	Deep learning
Rf regressor	Random forest regressor
DT regressor	Decision Tree regressor
LR	Linear Regression
KNN Regressor	K-Nearest Neighbors Regressor
XGBR	XGBoost Regressor
ADR	AdaBoost Regressor
IOT	Internet of Things

ACKNOWLEDGMENT

The author(s) sincerely thank the Doctors and doctoral students, Department of mathematics and computer science, Faculty of sciences ben msik, University of Hassan II, Casablanca, Morocco.

The author(s) appreciate the assistance of the staff of the Hassan II University, Casablanca, Morocco for supporting.

Special thanks go to Mme. Soumaya Ounacer and M. Elghoumari Mohammed Yassine and M. Azzouazi Mohamed.

REFERENCES

[1] Alrowais, Fadwa, Mashael M Asiri, Rana Alabdan, Radwa Marzouk, Anwer Mustafa Hilal, Ahmed Alkhayyat, et Deepak Gupta. « Hybrid Leader Based Optimization with Deep Learning Driven Weed Detection on Internet of Things Enabled Smart Agriculture Environment ». Computers and Electrical Engineering 104 (décembre 2022): 108411. https://doi.org/10.1016/j.compeleceng.2022.108411.

[2] Amrani, Abderraouf, Ferdous Sohel, Dean Diepeveen, David Murray, et Michael G.K. Jones. « Deep Learning-Based Detection of Aphid Colonies on Plants from a Reconstructed Brassica Image Dataset ». Computers and Electronics in Agriculture 205 (février 2023): 107587. https://doi.org/10.1016/j.compag.2022.107587.

[3] Chen, Wei, Yameng Xu, Zhe Zhang, Lan Yang, Xubin Pan, et Zhe Jia. « Mapping Agricultural Plastic Greenhouses Using Google Earth Images and Deep Learning ». Computers and Electronics in Agriculture 191 (décembre 2021): 106552. https://doi.org/10.1016/j.compag.2021.106552.

[4] Wang, X., & Zhang, Y. (2024). Data-Driven Approaches to Improve Agricultural Yield. *Journal of Data Science in Agriculture*, 15(3), 145-160.

[5] Gupta, D., & Khan, A. (2023). Machine Learning for Yield Prediction: A Review. *Computers and Electronics in Agriculture*, 200, 107-120.

[6] White, L., & Chen, Y. (2023). Machine Learning in Agriculture: A Review of Applications and Future Prospects. *Journal of Agricultural Sciences*, 15(4), 234-250.

[7] Liu, J., & Wang, Y. (2022). A Survey on the Applications of Deep Learning in Agriculture. *Computers and Electronics in Agriculture*, 185, 106-123.

[8] Patel, S., & Kumar, A. (2024). Advanced Data Analytics Techniques for Smart Agriculture: A Comprehensive Survey. *International Journal of Agricultural Science*, 12(2), 145-162.

[9] Quinlan, J. R. (1986). *Induction of Decision Trees*. Machine Learning, 1(1), 81–106.

[10] Breiman, L. (2001). *Random forests*. Machine Learning, 45(1), 5–32.

[11] Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis*. Wiley.

[12] Cover, T., & Hart, P. (1967). *Nearest neighbor pattern classification*. IEEE Transactions on Information Theory, 13(1), 21–27.

[13] Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

[14] Freund, Y., & Schapire, R. E. (1997). *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*. Journal of Computer and System Sciences, 55(1), 119–139.

[15] Cortes, C., & Vapnik, V. (1995). *Support-Vector Networks*. Machine Learning, 20(3), 273–297.

[16] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning representations by back-propagating errors*. Nature, 323(6088), 533–536.

[17] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). *Gradient-based learning applied to document recognition*. Proceedings of the IEEE, 86(11), 2278–2324.

[18] He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.

[19] Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences, 55(1), 119–139.

[20] Draper, N. R., & Smith, H. (1998). Applied Regression Analysis. Wiley.

[21] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

[22] Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. Machine Learning, 20(3), 273–297.

Quinlan, J. R. (1986). Induction of Decision Trees. Machine Learning, 1(1), 81–106.

[23] Patel, S., & Kumar, A. (2024). Advanced Data Analytics Techniques for Smart Agriculture: A Comprehensive Survey. International Journal of Agricultural Science, 12(2), 145-162.

[24] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794.

APPENDICES

Appendix 1. Comparative Study

Table I show the different results obtained in the articles and the different databases used and general context, also the different techniques used.

Appendix 2. Correlation Between Attributes

Table II describe the correlation between various agricultural attributes through a correlation matrix.

Appendix 3. Evaluation of Machine Learning Models

Table III describe the different results obtained in our study and the different algorithms used.

Appendix 4. Number of Productions by Season

Fig. 1 shows the number of agricultural productions by season.

Appendix 5. Temperature by season

Fig. 2 illustrates the temperature variations by season, showing that autumn has the highest average temperature, while winter exhibits the lowest.

K-Means and Morphology Based Feature Element Extraction Technique for Clothing Patterns and Lines

Xiaojia Ding[1]
Zhengzhou Academy of Fine Arts, Zhengzhou, 451450, China[1]

***Abstract*—In the process of clothing design and production, the traditional artificial feature element extraction method has the problems of low efficiency and insufficient precision, which is difficult to meet the automation and intelligent needs of modern clothing industry. In order to solve this problem, this paper proposes a technology that combines K-means clustering algorithm and morphology method to extract clothing pattern and line feature elements. This technology uses K-means clustering algorithm to preprocess clothing images to realize feature extraction of clothing pattern elements, and then introduces morphology method to realize feature extraction of image line elements. This technology not only improves the accuracy and efficiency of feature element extraction, but also retains the details of clothing images, which provides a strong support for automatic and intelligent processing in clothing design and production.**

Keywords*—K-means; morphological algorithm; feature extraction

I. Introduction

In the digital age, the field of clothing design and manufacturing is undergoing unprecedented changes. With the rapid development of computer vision, machine learning and image processing technology, automatic extraction and recognition of clothing pattern and line feature elements has become one of the key technologies to improve design efficiency and optimize production process. In 2017, Zhang Peng [1] compared and analyzed the applications of K-Dimensional tree's two nearest neighbor matching algorithm and random sampling agreement algorithm (RANSAC) in the field of image matching retrieval based on scale invariant feature transform (SIFT) algorithm. The results show that although K-Dimensional tree's NNN algorithm is superior in applicability and better in retrieval, RANSAC algorithm is superior in matching accuracy. In 2020, Jia Xiaojun [2] and his team successfully extracted feature elements from blue printed cloth through a series of image processing techniques, including gray level processing, corrosion, expansion, binarization and contour extraction, and built a database containing 12 elements accordingly, with an accuracy rate of 98.66%. At the same time, Chen Kaiqing [3] et al. deeply explored human preference feature extraction methods by using evolutionary matrix functions based on evolutionary strategies, and further studied how to combine deep learning with evolutionary strategies.

However, despite the remarkable progress in this field, there are still many problems. Traditional extraction methods are often inefficient and difficult to accurately capture the complex features of clothing patterns and lines, so it is particularly important to explore more efficient and accurate extraction techniques. K-means clustering algorithm and morphological analysis, as two powerful tools, have shown great potential and application value in the extraction of clothing feature elements. As a classical unsupervised learning method, K-means clustering algorithm has been widely used in many fields such as data analysis and image processing because of its simplicity and high efficiency. The algorithm divides the data into K clusters iteratively, so that the similarity of data points within the same cluster is high and the similarity of data points between different clusters is low [4]. In the feature extraction of clothing patterns, K-means algorithm can effectively group pixels or feature points in images according to their similarity, so as to identify the basic units of patterns and the direction of lines, providing strong support for subsequent design analysis and optimization. Morphological analysis is a shape-based image processing technology, which extracts the key features of the image by describing and analyzing the shape and structure in the image. In line feature extraction, morphological analysis can optimize and enhance the extracted feature elements through morphological operations, such as corrosion, expansion, open operation, close operation, etc., so that it is more in line with the design requirements and production requirements. This paper aims to explore the extraction technology of clothing pattern and line feature elements based on K-means clustering algorithm and morphological analysis, and demonstrate the application potential and practical effects of this technology in the field of clothing design and manufacturing through theoretical analysis and experimental verification [5].

Firstly, this paper gives a brief overview of K-means clustering algorithm and morphology algorithm, and then deeply discusses the extraction technology of clothing pattern and line feature elements based on these two algorithms. Specifically, the technical framework covers the K-means algorithm to extract pattern features from clothing images and the morphological algorithm to extract line feature elements in images. In order to verify the effectiveness and superiority of this comprehensive technology, a set of targeted experiments were designed and implemented in the end, and image data including clothing patterns were selected for testing.

II. K-means Algorithm and Morphology Algorithm

A. K-means Algorithm

K-means algorithm, also known as K-means clustering algorithm, is a widely used partitioning based clustering technology. Its core lies in the pre-set k value, which directly determines the number of initial cluster centers and thus

affects the partitioning of the whole data set. The K-means algorithm iteratively optimizes a cluster objective function, aiming to minimize the similarity between different clusters, while the data points within the same cluster show the greatest similarity [6]. The flexibility and effectiveness of K-means algorithm make it widely used in many fields. In the field of market segmentation, the algorithm can automatically divide the customer base into multiple market segments with different characteristics according to the customer's purchasing behavior or preference pattern, helping enterprises to develop more accurate marketing strategies. In the field of image processing, K-means algorithm is used in image segmentation tasks, which can realize automatic analysis and understanding of images by dividing pixels in images into different regions according to color, texture and other features, providing strong support for image recognition, image compression and other applications. In addition, K-means algorithm also plays an important role in text clustering, data mining, bioinformatics and other fields [7].

B. Morphology Algorithm

Mathematical Morphology is an algorithm system pioneered by French scholars J. Serra and Marceron in 1964, and has been deeply applied to many frontier fields such as computer vision, pattern recognition, image analysis and processing. As a powerful tool, it is widely used to analyze all kinds of complex images, such as the accurate identification of ground buildings in high-resolution satellite images from a high altitude, and the in-depth exploration of the structural mysteries of cells, bacteria, viruses and even lizard red blood cells in the biological microscopic world. Morphology has a series of basic operations, including expansion, corrosion, open operation and close operation. These operations not only play a significant role independently, but also can be skillfully combined and derived into a variety of practical algorithms such as top cap and bottom cap transformation, hit and miss transformation, watershed transformation, morphological gradient and particle analysis, which greatly enrich the means of image processing. Make image analysis more accurate and efficient [8]. As a core concept in morphology, the design and application of Structure Element directly determine the effect and efficiency of morphological processing. When applying mathematical morphology algorithm to process images, selecting the appropriate structural element is a crucial step, which determines how the algorithm "senses" and "transforms" the morphological structure in the image. With the continuous progress of technology and in-depth research, the application prospect of mathematical morphology will be broader, bringing more innovations and breakthroughs to the field of image analysis and processing [9].

III. Extraction of Clothing Pattern and Line Feature Elements Based on K-Means and Morphology

A. Clothing Pattern Extraction Based on K-means Algorithm

1) Cluster analysis: Cluster analysis aims to divide a large data set into multiple categories, ensuring that similar objects are highly similar and different categories are significantly different. This analysis technology has been widely used in many fields such as data compression, image segmentation, text retrieval and bioinformatics. Over the years, the research of cluster analysis has spawned many classical theories and algorithms, including hierarchical clustering, model clustering, grid clustering and partition clustering, etc. Each algorithm has its own characteristics and limitations. Hierarchical clustering can realize data grouping by building a tree structure, and the hierarchy is clear but difficult to trace back [10]. Model clustering depends on the strong robustness of the model set by the researcher; Grid clustering is efficient in spatial data processing and suitable for incremental implementation and high dimensional processing. Partitioning clustering is based on the distance between classes and is often limited to finding round or spherical clusters [11].

Clustering is significantly different from classification in that classification requires predictive training sets to extract classification rules and then apply them to unknown data, which is supervised pattern recognition. Cluster analysis, however, does not have such prior knowledge and automatically divides data into various categories by function or statistical method, which belongs to unsupervised pattern recognition.

The basic flow of cluster analysis covers the following key steps:

a) Data preprocessing: The core of this stage lies in the standardized data processing and feature dimension reduction. Standardization aims to eliminate the influence of different dimensions on the clustering results and ensure that each feature has the same weight in the clustering process. Feature dimensionality reduction reduces the dimensionality of the data set through methods such as principal component analysis (PCA) [12], while retaining key information in the original data as much as possible to improve the clustering efficiency and effect.

b) Feature screening: The process of identifying and selecting the most critical features in the original feature set helps reduce noise and redundant information, and makes the clustering process more focused on the essential characteristics of the data. The filtered features are converted into vector form, which is convenient for subsequent processing [13].

c) Feature refining: The selected features are further processed and extracted to obtain a more refined and effective feature representation for cluster analysis [14]. This step may involve advanced techniques such as feature transformation and feature fusion to improve the accuracy and stability of clustering.

d) Formulation of clustering strategies: According to prior knowledge, actual demand and data characteristics, the appropriate clustering criterion function is selected, which will guide the whole clustering process and ensure that the clustering results meet the expectations[15]. Under the guidance of criterion function, combined with mathematical tools and statistical theory, we perform clustering algorithm to divide data into different categories.

e) Evaluation of clustering results: After the clustering is completed, the quality of the clustering results needs to be evaluated, and the evaluation criteria mainly include intra-class validity evaluation, inter-class validity evaluation and correlation test [16]. Through comprehensive evaluation, we can judge the advantages and disadvantages of clustering results, and provide support for subsequent data analysis and decision making.

2) K-means clustering algorithm analysis: In this study, the K-means clustering algorithm is used to extract pattern feature elements from clothing images [17]. The algorithm aims to effectively divide n pixels in the image into K clusters, where k represents the number of preset color clusters. According to its RGB color value, each pixel is assigned to the cluster with the cluster center with the smallest Euclidean distance. The Euclidean distance is calculated as shown in Eq. (1).

$$d(X,C_i)=\sqrt{\sum_{j=1}^{m}(X_j-C_{ij})^2} \tag{1}$$

In formula 1, X is the data object; C_i is the i th cluster center. m is the dimension of the data object; X_j, C_{ij} is the JTH attribute value of X and C_i.

In the execution process, the distance between each pixel in the image and all clustering centers is first calculated, and each pixel is classified to the nearest clustering center based on this distance [18], and then the clustering center of each color cluster is iteratively updated by calculating the average value of RGB values of all pixels in the cluster as the new clustering center. This process is repeated until the convergence condition is met, that is, the location of the cluster center remains basically unchanged in successive iterations, which marks the completion of the clustering process. The algorithm flow chart is shown in Fig. 1.

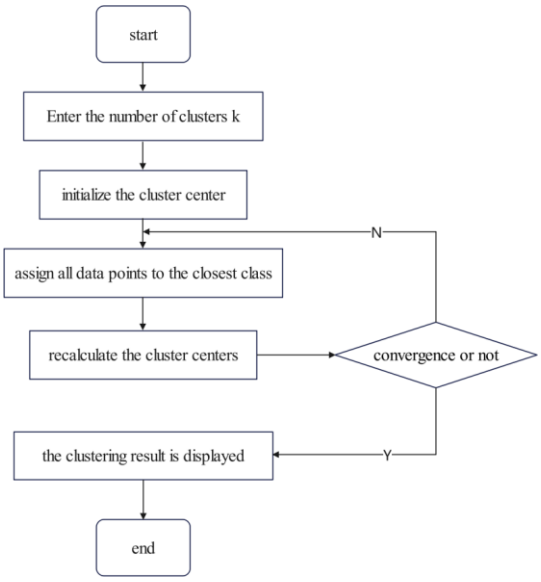


Fig. 1. K-means algorithm implementation flow.

The specific implementation steps of K-means clustering algorithm are as follows:

a) Initialization phase: Determine the number of clusters k and prepare a dataset with n data points $X=\{x_1, x_2, x_3,... x_n\}$. These data points can be pixels in an image, whose properties are RGB color values.

b) Select the initial clustering center: k data points are randomly selected from data set X as the initial clustering center, denoted as $m_1, m_2,... , m_k$ [19]. Each data point is then assigned to the nearest cluster center based on its distance to these cluster centers, resulting in k initial clusters $C=\{c_1, c_2,... , c_k\}$.

c) Calculate and update the cluster center: For each cluster c_i (i=1,2,... ,k), calculates the mean of the RGB color values of all data points in the cluster and uses this mean as the new cluster center m_i. The purpose of this step is to allow the cluster center to more accurately reflect the characteristics of the data points within its cluster.

d) Reassign data points to clustering: Iterate over each data point in the dataset again, calculating their distance to each new cluster center and reassigning each data point to the cluster to which the nearest cluster center belongs.

e) Iterative optimization: Repeat steps 3 and 4, that is, continuously compute new clustering centers and redistribute data points until a stop condition is met [20]. The commonly used stopping condition is that the change of cluster center is very small or the Sum of Squared Errors SSE (Sum of Squared Errors) no longer changes significantly, which indicates that the clustering results have become stable.

f) Output clustering results: The algorithm ends when the stop condition is reached, at which time k clusters are obtained, and each cluster contains a series of data points with high similarity in color features. The final output of this k subset is as clustering results, which can be used for subsequent pattern analysis, recognition, or other related applications.

The calculation formula of the squared error and SSE of the whole data set is Eq. (2).

$$SSE=\sum_{i=1}^{k}\sum_{X\in C_i}\left|d(X,C_i)\right|^2 \tag{2}$$

In Eq. (2), SSE represents the quality of clustering results; k is the number of clusters.

B. Garment Line Feature Element Extraction Based on Morphological Algorithm

In this study, morphological algorithm is used to accurately extract line feature elements in clothing. In the specific implementation process, corrosion and expansion operations in morphological processing are used to realize efficient identification and extraction of clothing line features [21].

Erosion is an image processing technique that removes pixels from the edges of the image so that the edges of the

image shrink inward. This process can not only fine-tune the edge shape of the image, but also effectively eliminate those objects in the original image whose size is smaller than a specific structural element, so as to remove noise and smooth the image edge. Corrosion operations are represented by a specific operator, denoted Θ. When A corrosion operation is applied to an image A to be processed, the selected structural element is S, and the operation can be expressed as $A \Theta S$, meaning that S is used to corrode A. The concrete implementation of corrosion operation is shown in Eq. (3).

$$A \Theta S = \{x : S + x \subset A\} \tag{3}$$

The main purpose of image processing by etching operation is to make the image boundary to be processed to shrink inward evenly, this shrinkage effect is usually manifested as the overall image shrinking a circle, and the specific size of this circle, that is, the degree of image shrinking inward, is directly affected by the shape and size of the selected structural elements [22]. As a key parameter in the corrosion operation, the structure element determines which boundary pixels will be removed, thus affecting the image shape after corrosion. Fig. 2 below shows the simulation effect of the corrosion operation, which intuitively shows the phenomenon of the inward contraction of the boundary after the image has been corroded.

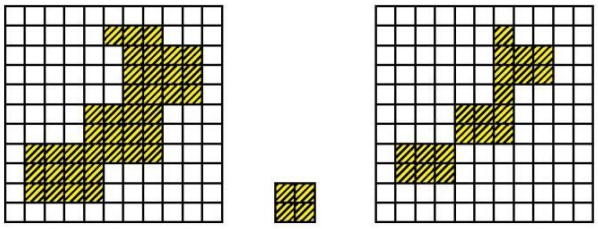


Fig. 2. Corrosion operation simulation diagram.

The principle of Dilation is the opposite of corrosion, it is an operation to expand the boundary of the target image by increasing the pixel points of the image boundary. The main function of the expansion operation is to connect the original discontinuous and breakpoint boundaries in the image, so as to restore or form a complete and smooth image boundary [23]. This is because the breakpoint on the image boundary may interfere or affect the subsequent computational statistical operations on the complete area of the image, and the expansion operation can effectively solve this problem. In expansion operations, the operator used is usually expressed as $\oplus$. When the expansion operation is performed on an image A to be processed, the selected structural element is S, and the operation can be expressed as $A \oplus S$, that is, S is used to expand A. The concrete implementation of the expansion operation is shown in Formula 4, which defines how to extend the boundary of image A based on the shape and size of the structural element S.

$$A \oplus S = \{a | A + S \cup A \neq \phi\} \tag{4}$$

In the dilation operation, an important property is that it satisfies the exchange rate of the operation, which means that the result of inflating image A with the structural element S (denoted as $A \oplus S$) is the same as that of "inflating" image A with the structural element S, as shown in Equation 5. After expansion processing, the boundary of the image to be processed will expand outward evenly for a circle, and the specific size of this circle is also affected by the shape and size of the selected structural elements [24]. Dilation operation can not only help to connect the original discontinuous boundaries in the image, but also fill the small holes or missing parts in the image to a certain extent, making the overall shape of the image more complete and coherent. Fig. 3 below shows the simulation effect of the expansion operation, and intuitively shows the phenomenon of the image boundary expanding outward after the expansion processing.

$$A \oplus S = S \oplus A \tag{5}$$

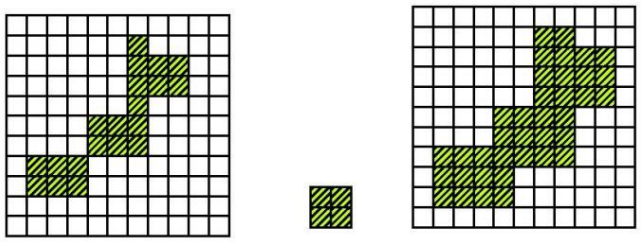


Fig. 3. Simulation diagram of expansion operation.

The corrosion operation effectively reduces the area of the target region by making the image boundary shrink inward, the essence of which lies in the removal of image edge pixels. On the contrary, the expansion operation increases the coverage of the target region by extending the image boundary outwards, making the object in the image look more expanded [25]. It is worth noting that whether it is corrosion or expansion, their changes to the image are mainly concentrated in the edge area, and the impact on the interior of the image is relatively small. Based on this property, when we subtract the corroded image from the original image, we can highlight the edge pixels that disappear due to corrosion and recover due to expansion, so as to effectively extract the edge outline of the object in the image.

The algorithms for edge detection using morphology generally include the following:

1) BM edge detection optimization technology: The core of BM edge detection technology is to reduce the noise effect by preprocessing. The strategy applies fuzzy filtering technology to reduce the noise in the image, and then performs the basic operation of morphology. In this process, the width of the carefully designed structural element B matches the size of the local average processing window, aiming to capture the edge information more accurately and reduce the noise interference at the same time, which is expressed as Eq. (6).

$$BM(f) = \min\{f_{av} - (f_{av} \Theta B), (f_{av} \oplus B) - f_{av}\} \tag{6}$$

In Formula 6, B represents the structural element to be used, and f_{av} represents the gray value of the image after fuzzy processing.

2) ATM edge detection strategy: ATM edge detection method introduces α-adjustment mechanism, the core of which is to optimize the input image by adjusting parameter α, so as to effectively suppress the noise in the image. This α-adjustment strategy provides a clearer and less noisy input environment for image edge detection, thus improving the accuracy and reliability of edge detection. The expression of α-adjustment is Eq. (7).

$$f = \sum_{l=\alpha+1}^{k-\alpha} \frac{f(l)}{k-2\alpha} \tag{7}$$

The formula expression of edge detection is shown in equation 8.

$$ATM(f) = \min\{(f \bullet B)-(f\Theta B),(f \oplus B)-(f \circ B)\} \tag{8}$$

3) ASF multi-directional edge detection: ASF edge detection method is a multi-dimensional analysis method. It adopts the structural elements in the four directions of 0°, 45°, 90° and 135° with the X-axis, and performs the combination of open γ_l operation and close φ_l operation on the image to capture the edge features in different directions, and then implements corrosion operations on the processed image to further refine and strengthen the edge information. The formula is shown in Eq. (9).

$$f = (\gamma_l \varphi_l \Theta B) - \gamma_l \varphi_l \tag{9}$$

In the field of morphological edge detection, the size and shape of structural elements are directly related to the detection effect. Larger structural elements can effectively suppress noise, but may be at the expense of edge details, while smaller structural elements can capture edge features more finely, but may not completely eliminate noise. Therefore, the ideal structural element should be able to remove noise while retaining the edge details. The directivity of the edge is also significantly affected by the shape of the structural element, with different shapes having higher sensitivity to edges of a particular shape. In order to overcome this limitation and improve the anti-noise ability of edge detection, an effective strategy is to use multi-structural elements combined with multi-scale morphological edge detection operators. By combining structural elements of different sizes and shapes, this method can capture edge information in images more comprehensively and enhance the robustness to noise, so as to achieve high-precision image edge detection. In practical applications, in order to obtain detailed elements such as clothing line outline more accurately, the image will be preprocessed first, including noise elimination to reduce interference, binary processing to simplify image information and possibly reverse the light and dark areas in the image by taking the inverse operation, and then combined with expansion and corrosion operations to further emphasize and extract the clothing line outline elements.

IV. EXPERIMENT

This experiment aims to extract key pattern and line feature elements from a set of image data containing clothing patterns by K-means clustering algorithm. The main objective of the experiment is to verify the practicability and effectiveness of K-means algorithm in image processing and feature extraction tasks, and to explore its potential value in practical application scenarios such as clothing design and automatic image classification and recognition.

The experiment first imports the image dataset into a Python environment, and performs a series of pre-processing steps on these images, including but not limited to conversion to grayscale, resizing the image to ensure consistency, and normalization of pixel values. After feature processing, the image data is converted into a format suitable for K-means algorithm. The K-means clustering algorithm is applied to perform cluster analysis on these feature vectors, and similar pattern and line elements are grouped into different clusters through the iterative process. After the clustering is completed, visualization technology is used to visually present the clustering results, as shown in Table I.

TABLE I. RESULTS OF PATTERN LINE EXTRACTION

Master drawing	
Pattern extraction	
Line extraction	

After the experimental application of K-means clustering algorithm to cluster the feature vectors of image data, the experiment successfully groups similar pattern and line elements into different clusters, which shows that K-means algorithm has strong effectiveness and accuracy in image feature extraction and grouping. At the same time,

visualization technology is used to visually present the clustering results, and the differences and similarities between different clusters can be clearly seen from Table I, which further verifies the effectiveness of K-means clustering algorithm in image feature extraction and grouping.

Experiments show that in the complex and changeable background environment, K-means algorithm not only has remarkable clustering effect, but also has strong anti-noise property, ensuring the accuracy of segmentation results. However, the automation degree of this algorithm is still insufficient, and its performance is highly dependent on the setting of key parameters by experienced operators. Meanwhile, the long operation time also limits its application potential in the pursuit of efficient processing scenarios. In contrast, morphology algorithm stands out for its short operation time and high processing efficiency, but its effect is easily affected when processing images containing noise, especially when processing color images, complexity and timeliness become challenges. Nevertheless, the advantages of morphology algorithm in spatial and frequency domain image recognition cannot be ignored. It can effectively retain the original information of the image while filtering the noise, which lays a solid foundation for the subsequent image processing. The limitations of a single algorithm in extraction accuracy, efficiency or specific target recognition are obvious, so exploring a new path of algorithm fusion has become the key to improve the ability of image feature extraction. By organically combining and optimizing different algorithms to build a multi-level recognition and extraction system, more accurate and efficient image feature extraction can be achieved. This fusion strategy can not only make full use of the advantages of each algorithm, but also make up for each other's shortcomings, so as to improve the overall performance. In addition, with the rapid development of deep learning technology, its application in the field of pattern segmentation and extraction undoubtedly opens up a new way to improve the extraction accuracy and automation level. Deep learning algorithms have strong feature extraction and pattern recognition capabilities, and can automatically learn and adapt to complex image features, which also puts higher requirements on the technical requirements of computing equipment. More computing power is needed to support the operation of complex algorithms and programs. Nevertheless, I think this challenge is worthwhile, because the introduction of deep learning technology will greatly promote the development of clothing pattern and line feature element extraction technology, and bring more intelligent and efficient solutions to the field of clothing design and manufacturing.

V. Conclusion

This paper deeply discusses the extraction technology of clothing pattern and line feature elements based on K-means clustering algorithm and morphology method, expounds the basic principle of K-means algorithm and morphology algorithm in detail, and focuses on how to use K-means algorithm to perform pixel-level clustering analysis of clothing images. The morphology algorithm is introduced to achieve fine extraction of line feature elements. This paper describes the implementation process of the whole feature extraction technology in detail, and finally verifies the effectiveness and superiority of K-means and morphology algorithm in the extraction of clothing pattern and line feature elements through experiments. The experimental results show that this technology can accurately and efficiently identify and extract key patterns and line features in clothing images. It provides powerful technical support for clothing design, analysis, retrieval and other fields. Looking forward to the future, with the continuous development of artificial intelligence technology and the continuous optimization of algorithms, the extraction technology of clothing pattern and line feature elements based on K-means and morphology will have a broader application prospect. It can help fashion designers quickly capture trends, innovative design ideas, but also to provide strong support for the automation and intelligence of the clothing production process, while the technology is also expected to play an important role in personalized customization, virtual reality fitting, intelligent collocation recommendation and other fields, to promote the entire fashion industry to more efficient, intelligent, personalized direction.

References

[1] Zhang P. Research on the application of SVM in Beijing Opera facial mask image recognition [D/OL]. Beijing: Beijing University of Technology, 2017.

[2] JIA X J, LIU Z H.Element extraction and convolutional neural network-based classification for blue calico [J]. Textile research journal,2020,1-17.

[3] CHEN K Q, HUANG X Q.Feature extraction method of 3D art creation based on deep learning[J].Soft computing,2020,24(11):8149-8161.

[4] Jain P ,Tiwari A ,Som T .Intuitionistic fuzzy rough set model based on k-means and its application to enhance prediction of aptamer–protein interacting pairs.Journal of Ambient Intelligence and Humanized Computing,2024,15(9):3575-3586.

[5] Pan C .Construction of Risk Prediction Models for Enterprise Finance Sharing Operations Using K-Means and C4.5 Algorithms.International Journal of Computational Intelligence Systems, 2024, 17(1):208-208.

[6] Folly M ,Amaral C L ,Silveira D L F L , et al.Gastrotheca Fitzinger, 1843 tadpole morphology: Larval cranium description and its evolutionary implications (Amphibia: Anura: Hemiphractidae)..Journal of morphology, 2024, 285 (9):e21766.

[7] Fang K ,Hou Y ,Niu L , et al. Individualized gray matter morphological abnormalities uncover two robust transdiagnostic biotypes. Journal of Affective Disorders,2024,365193-204.

[8] Chajaei F , Bagheri H .Machine learning framework for high-resolution air temperature downscaling using LiDAR-derived urban morphological features.Urban Climate,2024,57102102-102102.

[9] Chen L ,Roe R D ,Kochert M , et al.k-Means NANI: An Improved Clustering Algorithm for Molecular Dynamics Simulations..Journal of chemical theory and computation,2024,

[10] Sofi A S ,Rafiq S ,Majid D , et al.Irradiation-induced modifications of apple seed protein isolate: Exploring techno-functional, structural, thermal, and morphological characteristics.Radiation Physics and Chemistry,2024,225112139-112139.

[11] Bourmaud A ,Rudolph L A ,Habrant A , et al.Innovative use of fluorescent grafting for damage and morphological monitoring during the extrusion process of PCL-flax fibre-reinforced composites.Industrial Crops & Products,2024,222(P1):119446-119446.

[12] Hazendonk V S L ,Tuinier R ,Foschino E , et al.Morphological analysis of polydisperse nanoplatelets using SAXS.Colloids and Surfaces A: Physicochemical and Engineering Aspects,2024,702(P1):134997-134997.

[13] Choi J S ,Jiao J .Uncovering electric vehicle ownership disparities using K-means clustering analysis: A case study of Austin, Texas.Journal of Computational Social Science,2024,(prepublish):1-54.

[14] Ji R ,Yang J ,Wu Y , et al.Correction to: Construction and analysis of students' physical health portrait based on principal component analysis improved Canopy-K-means algorithm.The Journal of Supercomputing,2024,80(14):21561-21562.

[15] Sarah P ,Krishnapriya S ,Saladi S , et al.A novel approach to brain tumor detection using K-Means++, SGLDM, ResNet50, and synthetic data augmentation.Frontiers in Physiology,2024,151342572-1342572.

[16] Zhang X ,Li F ,Shi Z , et al.Multi-view reduced dimensionality K-means clustering with [formula omitted]-norm and Schatten [formula omitted]-norm.Pattern Recognition,2024,155110675-110675.

[17] Vardakas G ,Likas A .Global k-means++: an effective relaxation of the global k-means clustering algorithm.Applied Intelligence,2024,54(19):8876-8888.

[18] Chen S ,Ouyang X ,Luo F .Ensemble One-Class Support Vector Machine for Sea Surface Target Detection Based on k-Means Clustering.Remote Sensing,2024,16(13):2401-2401.

[19] Meng J ,Abed M A ,Elsehrawy G M , et al.Nano-integrating green and low-carbon concepts into ideological and political education in higher education institutions through K-means clustering.Heliyon,2024,10(10):e31244-.

[20] Sun L ,Chen Y ,Du Q , et al.Identification of low-voltage phase lines using IEC 61850 and K-means clustering.Electric Power Systems Research,2024,234110597-.

[21] Jiang H ,Xu W ,Chen W , et al.Performance of the digital cell morphology analyzer MC-100i in a multicenter study in tertiary hospitals in China..Clinica chimica acta; international journal of clinical chemistry,2024,555117801-117801.

[22] Ye X ,Fang L ,Chen Y , et al.Performance comparison of two automated digital morphology analyzers for leukocyte differential in patients with malignant hematological diseases: Mindray MC-80 and Sysmex DI-60..International journal of laboratory hematology,2024,46(3):457-465.

[23] Shariq M S ,Ansar Z A .Digital Morphology: Bridging the Final Gap in Automated Haematology Testing..Journal of the College of Physicians and Surgeons--Pakistan : JCPSP,2023,33(8):949-949.

[24] Wei J ,Zaibin C .Defect detection on solar cells using mathematical morphology and fuzzy logic techniques.Journal of Optics,2023,53(1):249-259.

[25] Oğuzhan E A ,Gülşen T .An earthquake damage identification approach from VHR image using mathematical morphology and machine learning.Neural Computing and Applications,2022,34(21):18757-18771.

Simulation Analysis of Obstacle Crossing Stability for Transmission Line Inspection Robot

Qianli Wang
School of Mathematics and Computer Science, Yichun University, Yichun, 336000, China

***Abstract*—As an indispensable energy source in production and daily life, electricity has important implications in the operation of society and economic development. As the hub of power transmission, the safety of transmission lines is related to the stability of the power grid. Regular inspection of transmission lines is an effective measure to ensure the stability of the power system. Patrol robots are used for regular inspection of transmission lines due to their advantages such as low cost and long running time. To achieve collision free, the study proposes an obstacle path planning algorithm based on an improved bidirectional fast expanding random tree through kinematic analysis of the robot. According to the experimental results, when overtaking the damper, the rotation ranges of the 1st claw/arm and the 2nd bracket/arm/claw were (0°~22°), (-50°, 10°), (0°, 25°), (-50°, 10°), and (0°, 22°), respectively. The corresponding rotational speeds were (-1.5~1.5) deg/s, (-3, 2.5) deg/s, (-3.5~3.5) deg/s, (-2.5, 3) deg/s, and (-27, 2) deg/s, respectively. The expansion and contraction ranges of the upper, middle, lower, and horizontal push rods were (0, 100) mm, (0, 110) mm, (-60, 10) mm, and (0, 20) mm, respectively. From the above results, when crossing obstacles, the motion acceleration of the inspection robot is not significant. The speed changes smoothly. The obstacle crossing path planning algorithm proposed in the study can achieve stable motion of the inspection robot.**

Keywords*—Transmission line; inspection robot; obstacle crossing path; kinematic analysis; bidirectional fast expanding random tree

I. INTRODUCTION

To meet daily production and living needs, power plants transmit electricity to various regions through high-voltage transmission lines. Due to long-term exposure of high-voltage transmission lines to the wild, various faults continue to occur, posing a threat to the safety of the power grid. Meanwhile, with the continuous growth of electricity demand, the mileage of high-voltage transmission lines has also rapidly increased. A large number of high-voltage transmission lines penetrate into areas that are rarely visited, posing great difficulties for circuit inspection work. This makes it difficult to detect transmission line faults in a timely manner, increasing the probability of safety accidents occurring [1-2]. Compared with manual inspections and drone inspections, inspection robots have high inspection quality, low cost, and long working hours, making them suitable for various environments. In addition to cables, there are also various metal fittings on the transmission line, which are obstacles for inspection robots. Therefore, to ensure the inspection range, the inspection robot needs to cross these obstacles [3-4]. In recent years, with the vigorous development of various technologies such as machine vision, deep learning, intelligent sensors, and intelligent Internet of Things, overhead transmission line inspection robots have also ushered in new development opportunities. For example, combining machine vision, deep learning, and multi-sensor information fusion technologies can achieve online obstacle recognition, localization, and model construction. Based on obstacle models, online obstacle crossing path planning can be performed to ensure that the patrol robot has obstacle crossing capabilities in different obstacle and attitude environments. It can also be combined with intelligent IoT technology to achieve real-time feedback monitoring and adjustment of the entire obstacle crossing process. Therefore, the research on obstacle crossing path planning for overhead transmission line patrol robots in this article has great practical application significance. In the current field of 110kV overhead transmission line inspection robots, the main obstacle crossing technology solutions include linear guide rail and linkage robotic arm. However, each has its own advantages and disadvantages. The former is stable and reliable when crossing obstacles, but the overall weight is too large, the load capacity is low, and the obstacle crossing distance is short. The latter has lower overall weight and improved obstacle crossing distance, but requires higher requirements for rotary joint motors and lower reliability in obstacle crossing. Meanwhile, due to the complexity of obstacle crossing actions, it is difficult to ensure that the relative position and posture of the line patrol robot and obstacles are fixed during each obstacle crossing operation. Pre-planning and fixing the obstacle crossing action sequence for each obstacle can easily lead to collisions with obstacles, wires, etc. during the obstacle crossing process, resulting in failure. Therefore, to improve the work efficiency of Transmission Line Inspection Robot (TLIR) and achieve collision free motion of inspection robots, a cost-effective obstacle crossing path planning method for inspection robots based on the improved Rapidly Expanding Random Tree-Connect (RRT-Connect) is proposed. By improving joint motion and search step size, and setting an intermediate random tree, the RRT-Connect algorithm being prone to local search is improved, simplifying the search process.

Rest of the study is divided into five sections. Section II will briefly describe the research status of transmission line inspection robots and RRT algorithms. Section III will conduct dynamic analysis on the optional robot and designs the obstacle crossing path planning algorithm. Section IV will simulate and analyze the obstacle crossing performance of the inspection robot. Section V will discuss the research results to further analyze them. Section VI will summarize the research of the entire article.

II. Related Works

With the increasing demand for transmission lines in the power system, it has become a current trend for inspection robots to conduct line inspections. Wang et al. [5] proposed a new four arm inspection robot to address the low obstacle crossing efficiency and poor safety of inspection robots. This new inspection robot had a rectangular frame on its walking arm to improve obstacle crossing efficiency. In addition, the robot could also use two different sets of arms to cross the tension tower together. The experimental results showed that the new four arm inspection robot could cross common obstacles on transmission lines. Wei et al. [6] proposed a lateral friction recognition method based on fuzzy control to address the limited movement of inspection robots. This method monitored the friction between the walking wheel line and the wheel flange, effectively avoiding "wheel line" jamming caused by external uncertain factors. This method could effectively improve the automation and intelligence level of the operation and transmission system of the inspection robot. Huang et al. [7] proposed a detection method based on visual sensors for rapid detection of wire looseness issues. This method collected wire images through an image acquisition system. The GEG algorithm was used for foreground highlighting and hierarchical enhancement. The implementation results showed that this method could quickly detect loose wires. Qingkai et al. [8] proposed a fault detection method based on lightweight semantic segmentation network to address the accuracy in line fault detection. A two branch network was designed to locate instances and refine contours. This method improved the accuracy of line fault detection. Xiong et al. [9] proposed an object recognition model based on human concept learning to address the automatic fault detection for inspection robots. The model was preliminarily identified through Mask R-CNN. Bayesian contextual network was used for result correction. Compared with other methods, the accuracy of this method was increased by 9.7%.

As a tree-based algorithm, RRT is suitable for multi-dimensional space, with simple principles and strong applicability. It is a typical path planning algorithm. Hu B et al. [10] proposed a path planning algorithm based on RRT to find accessible paths for wheeled robots. This algorithm used a motion control law guided by a posture-based steering function to reach its destination. From the experimental results, the trajectory generated by this method was smoother and shorter in length. Sun et al. [11] proposed a motion planning method based on the SC-RRT algorithm for autonomous navigation of mobile robots. This method improved the branch growth mode of the RRT-Connect algorithm, reduced the sampling space, and quickly obtained a path away from obstacles. Jhang et al. [12] proposed a motion planning method based on bidirectional fast exploration random trees and Reeds-Shepp curves for the path planning problem of autonomous parking. This method could be independently proposed in complex parking lots. Zammit and Kampen [13] proposed a path planning algorithm based on RRT for real-time safe navigation of unmanned aerial vehicles. From the experimental results, in simple location environments, the RRT algorithm had a success rate of over 95%. Compared with other algorithms, the path was shorter. Wang et al. [14] proposed a hybrid bidirectional fast exploration random tree based on reinforcement learning to address the randomness and slow convergence speed in path generation. This method improved search efficiency and shortened the generation path.

In summary, the research on TLIR has been quite effective. However, most studies overlook the stability of the obstacle crossing process. The RRT algorithm, as a typical path planning algorithm, although suitable for multi-dimensional spaces, is prone to falling into local search. Therefore, to ensure the motion stability of the inspection robot, an obstacle crossing path planning algorithm based on the improved RRT-Connect algorithm is proposed. By improving joint motion and search step size, and setting an intermediate random tree, the traditional RRT-Connect is improved.

III. Obstacle Crossing Stability Analysis of TLIR

In transmission lines, in addition to equipment such as tension towers and cables, it also includes hardware such as insulator strings and spacer rods. These hardware tools that exist in transmission lines are obstacles for TLIR. During the inspection process, whether the robot can stably cross various obstacles is the key to complete the inspection work.

A. Dynamic Analysis of Obstacle Crossing Motion of TLIR

During the obstacle crossing process of the TLIR, one claw of the robot grasps the line and the other claw disengages. The motion of the detachment claw is similar to the reverse motion of the gripping claw. To analyze the obstacle crossing engineering of robots, a kinematic analysis is conducted on the obstacle crossing process of the inspection robot using the D-H parameter method with a horizontal plane as the reference plane. The joint change matrix of the inspection robot is shown in Eq. (1).

$$ {}_{i}^{i-1}T = \begin{bmatrix} \cos\theta_i & -\sin\theta_i\cos\alpha_i & \sin\theta_i\sin\alpha_i & a_i\cos\theta_i \\ \sin\theta & \cos\theta_i\cos\alpha_i & -\cos\theta_i\sin\alpha_i & a_i\sin\theta_i \\ 0 & \sin\alpha_i & \cos\alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1) $$

In Eq. (1), θ_i represents the amount of rotation of the rotating joint, (rad). d_i represents the extension and retraction of the arm, (mm). α_i and a_i are constants. In the link model of the TLIR, the angle relationship between the upper push rod, middle push rod, and wheel claw is shown in Fig. 1.

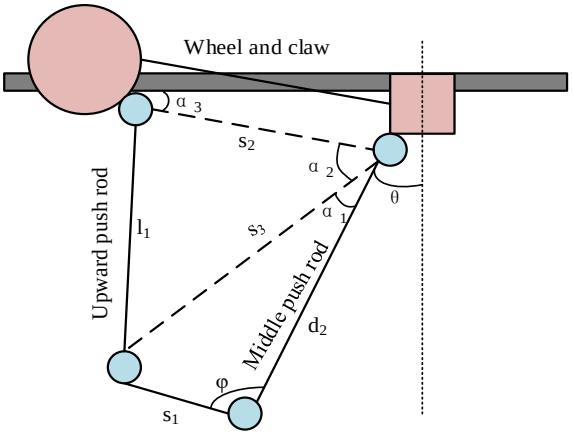


Fig. 1. Upper bar / middle bar and wheel claw angle relationship.

In Fig. 1, the upper push rod is responsible for adjusting the angle between the middle push rod and the wheel claw. According to the relationship shown in the above figure, the geometric relationship among the three is shown in Eq. (2).

$$\begin{cases} s_3 = \sqrt{s_1^2 + d_2^2 - 2s_1 d_2 c\cos\varphi} \\ \alpha_1 = \arccos\left(\dfrac{s_3^2 + d_2^2 - s_1^2}{2s_3 d_2}\right) \\ \alpha_2 = \arccos\left(\dfrac{s_3^2 + s_2^2 - l_1^2}{2s_3 s_2}\right) \\ \theta = \pi - \left(\dfrac{\pi}{2} - \alpha_3\right) - (\alpha_1 + \alpha_2) \end{cases} \quad (2)$$

In Eq. (2), s_3 represents the distance between joint 1 and joint 2. s_1 stands for the length of the connector. d_2 stands for the length of the middle push rod. φ stands for the angle between the middle push rod and the connecting piece. α_1 signifies the angle between s_3 and the middle push rod. α_2 stands for the angle between the wheel claw and s_3. l_1 represents the length of the upper push rod. θ refers to the angle between the middle push rod and the vertical line. α_3 represents the angle between the wheel claw and the horizontal line. According to the length of the middle push rod, the range of the angle between the upper push rod and the wheel claw can be determined. The angle relationship between the lower/middle push rod and the center turntable is illustrated in Fig. 2.

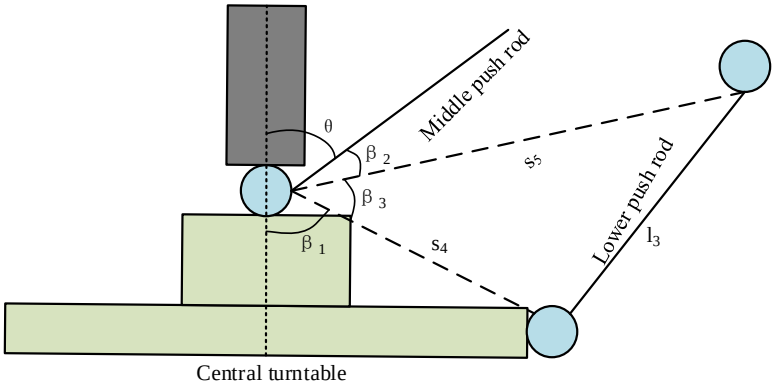


Fig. 2. Angle relationship between lower / middle push rod and center turntable.

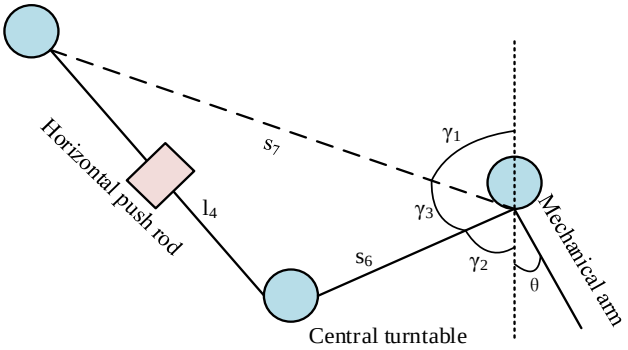


Fig. 3. Angle relationship between horizontal lever / robotic arm and center turntable.

In Fig. 2, the angle between the middle push rod and the center turntable is adjusted by the lower push rod. According to the relationship shown in Fig. 2, the geometric relationship among the three is shown in Eq. (3).

$$\begin{cases} \beta_2 = \arccos\left(\dfrac{s_4^2 + s_5^2 - l_3^2}{2s_4 s_5}\right) \\ \theta = \beta_1 + \beta_2 + \beta_3 - \pi \end{cases} \quad (3)$$

In Eq. (3), β_3 represents the angle between the center turntable and s_5. s_4 represents the length of the central turntable. s_5 represents the distance between joint 5 and the upper endpoint of the central turntable. l_3 refers to the length of the lower push rod. β_1 refers to the angle between the center turntable and the vertical line. β_2 is the angle between the middle push rod and s_5. The relationship between the horizontal push rod/robotic arm and the center turntable angle is shown in Fig. 3.

In Fig. 3, the angle between the robotic arm and the central turntable is adjusted by a horizontal push rod. According to the relationship shown in the above figure, the geometric relationship expression among the three is shown in Eq. (4).

$$\begin{cases} \gamma_3 = \arccos\left(\dfrac{s_6^2 + s_7^2 - l_4^2}{2 s_6 s_7}\right) \\ \theta' = \gamma_1 + \gamma_2 + \gamma_3 - \pi \end{cases} \tag{4}$$

In Eq. (4), γ_3 represents the angle between s_7 and the central turntable. s_6 represents the length of the central turntable. l_4 refers to the length of the horizontal push rod. θ' refers to the angle between the robotic arm and the vertical line. γ_1 refers to the angle between s_7 and the vertical line. γ_2 refers to the angle between the center turntable and the vertical line. The D-H parameters of each connecting rod can be combined with Eq. (1) to obtain the homogeneous transformation matrix of each joint. Directly below the transmission line is the origin of the global coordinate system. The homogeneous transformation matrix of the inspection robot end in the global coordinate system can be obtained by multiplying the Homogeneous Transformation Matrices (HTM) between different coordinate systems, as displayed in Eq. (5).

$${}_7^G T = {}_0^G T \cdot {}_1^0 T \cdot {}_2^1 T \cdot {}_3^2 T \cdot {}_4^3 T \cdot {}_5^4 T \cdot {}_6^5 T \cdot {}_7^6 T = \begin{bmatrix} n_x & o_x & a_x & p_x \\ n_y & o_y & a_y & p_y \\ n_z & o_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

In Eq. (5), ${}_7^G T$ stands for the HTM of coordinate system 7 relative to the global coordinate system. ${}_0^G T$ stands for the HTM of coordinate system 0 relative to the global coordinate system. ${}_1^0 T$, ${}_2^1 T$, ${}_3^2 T$, ${}_4^3 T$, ${}_5^4 T$, ${}_6^5 T$ and ${}_7^6 T$ represent the HTM of coordinate system 1/2/3/4/5/6/7 relative to coordinate system 0/1/2/3/4/5/6, respectively. To obtain the desired joint angle, the inverse kinematics analysis on the robot is performed. The inspection robot has seven joints and many variables, which are difficult to solve. The special structure of the inspection robot allows it to obtain the inverse kinematics expression without considering the last joint variable. At this time, the HTM of the global coordinate system at the end of the robot is shown in Eq. (6).

$${}_6^G T = {}_0^G T \cdot {}_1^0 T \cdot {}_2^1 T \cdot {}_3^2 T \cdot {}_4^3 T \cdot {}_5^4 T \cdot {}_6^5 T = \begin{bmatrix} n'_x & o'_x & a'_x & p'_x \\ n'_y & o'_y & a'_y & p'_y \\ n'_z & o'_z & a'_z & p'_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6}$$

When the rotation of the second robotic arm is not 0, the analytical expressions for the first six joint variables are shown in Eq. (7).

$$\begin{cases} \theta_1 = \arctan\left(d_2 \sin\theta_1 / d_2 \cos\theta_1\right) \\ d_2 = d_2 \cos\theta_1 / \cos\theta_1 \\ \theta_3 = \left(\theta_1 + \theta_3\right) - \theta_1 \end{cases} \tag{7}$$

In Eq. (7), θ_1 represents the rotational amount of the claw 1. d_2 represents the extension and retraction amount of arm 1. θ_3 represents the rotation amount of arm 1. When the rotation of the arm 2 is 0, the analytical expressions for the first six joint variables are shown in equation (8).

$$\begin{cases} \theta_1 = \arcsin o_x \\ d_2 = \dfrac{\dfrac{p_x - 266}{\sin\theta_1} + \left(\dfrac{p_z}{-\cos\theta_1}\right)}{2} \\ d_6 = \dfrac{\dfrac{p_x - 266}{\sin\theta_1} - \left(\dfrac{p_z}{-\cos\theta_1}\right)}{2} \end{cases} \tag{8}$$

In Eq. (8), d_6 represents the telescopic amount of arm 2. The length of each push rod is shown in Eq. (9).

$$\begin{cases} l_1 = \sqrt{s_2^2 + s_3^2 - 2 s_2 s_3 \cos\left(\dfrac{\pi}{2} - \alpha_1 + \alpha_3 - \theta_1\right)} \\ l_3 = \sqrt{s_4^2 + s_5^2 - 2 s_4 s_5 \cos\left(\pi - \beta_1 - \beta_2 + \theta_3\right)} \\ l_4 = \sqrt{s_6^2 + s_7^2 - 2 s_6 s_7 \cos\left(\pi - \gamma_1 - \gamma_2 + \theta_4\right)} \\ l_5 = \sqrt{s_4^2 + s_5^2 - 2 s_4 s_5 \cos\left(\pi - \beta_1 - \beta_2 + \theta_5\right)} \\ l_1 = \sqrt{s_2^2 + s_3^2 - 2 s_2 s_3 \cos\left(\dfrac{\pi}{2} - \alpha_1 + \alpha_3 - \theta_7\right)} \end{cases} \tag{9}$$

In Eq. (9), l_1 - l_7 respectively represent the length of the push rod that adjusts the corresponding joint variable.

B. Obstacle Crossing Path Planning Algorithm for TLIR

To ensure the collision free motion of the TLIR during the operation, an obstacle crossing path planning algorithm based on the improved RRT-Connect algorithm is proposed by combining its kinematic analysis model. RRT is a path planning algorithm based on random sampling, which can effectively overcome the high dimensionality caused by motion. The search tree extension diagram of the RRT is displayed in Fig. 4.

In Fig. 4, when using the initial point as the extended root node and conducting search expansion in space, each search randomly generates sampling points. If the distance between the random sampling point and its nearest point is greater than the set extension step, the nearest point will be extended to the random sampling point at the set step size, generating a new node. Otherwise, the random sampling points are used as new nodes [15-17]. If the connection between the new node and the nearest node collides with an obstacle, the new node will be abandoned. Otherwise, the new node is added to the random tree. The distance between nodes is shown in Eq. (10).

$$\rho\left(x_1, x_2\right) = \left\| x_2 - x_1 \right\| \tag{10}$$

In Eq. (10), x_1 and x_2 represent different nodes. The new node is shown in Eq. (11).

$$x_{new} = x_{near} + \lambda \frac{x_{rand} - x_{near}}{\left\| x_{rand} - x_{near} \right\|} \tag{11}$$

In Eq. (11), x_{new} represents the new node. x_{near} represents the node closest to the random sampling point. x_{rand} represents a random sampling point. λ represents the extension step size. The RRT algorithm has strong randomness, resulting in a large number of redundant nodes. Path planning takes a long time. The RRT-Connect algorithm improves the search speed by incorporating the Connect heuristic algorithm on the basis of bidirectional RRT algorithm, which has a certain degree of directionality. The new node generation method of RRT-Connect algorithm is shown in Eq. (12).

$$\begin{cases} x_{new} = x_{near}\left(a + \varepsilon\cos\theta, b + \varepsilon\sin\theta\right) \\ \theta = \arctan\left(x_{target}(b) - x_{near}(b), x_{target}(a) - x_{near}(a)\right) \end{cases} \tag{12}$$

In Eq. (12), (a,b) represents the node coordinates. θ represents the growth angle between the line connecting the target node and the nearest node and the horizontal line. ε represents the minimum fixed step size. x_{target} represents the target node. The RRT-Connect greatly improves the efficiency of path planning. However, the mapping relationship between the joint space and workspace of the TLIR is complex, resulting in significant differences in the pose changes between the new node and the parent node in the workspace. It is difficult to apply to the obstacle crossing path planning. Therefore, the RRT-Connect algorithm is improved by improving joint motion, search step size, and setting intermediate random trees. During robot collision detection, when there is a significant change in the end pose of the workspace between two nodes, interpolation collision detection needs to be performed between the two nodes. At this point, it is difficult to distinguish the impact of different joint movements on the changes in workspace posture. Therefore, to simplify the collision detection process, a path search method based on single joint motion is proposed, which means that the child nodes only change one joint during the expansion process [18-20]. Although the single joint motion path search method can achieve clear correspondence between joint space and workspace pose. However, in joint space, different single joint movements can still cause changes in the workspace posture. Therefore, to simplify the search process, the fixed position step size is set within the workspace instead of the joint space. At this point, collision detection only needs to detect the sampling points and nearest nodes, which greatly reduces the workload of joint space collision detection. In addition, the RRT-Connect algorithm is prone to getting stuck in local search when encountering large obstacles. A large number of random sampling points from other directions need to be generated to cross the obstacle. Therefore, to alleviate the local search problem, a random tree with an intermediate point is set in the middle of the path. The improved RRT-Connect path planning process is displayed in Fig. 5.

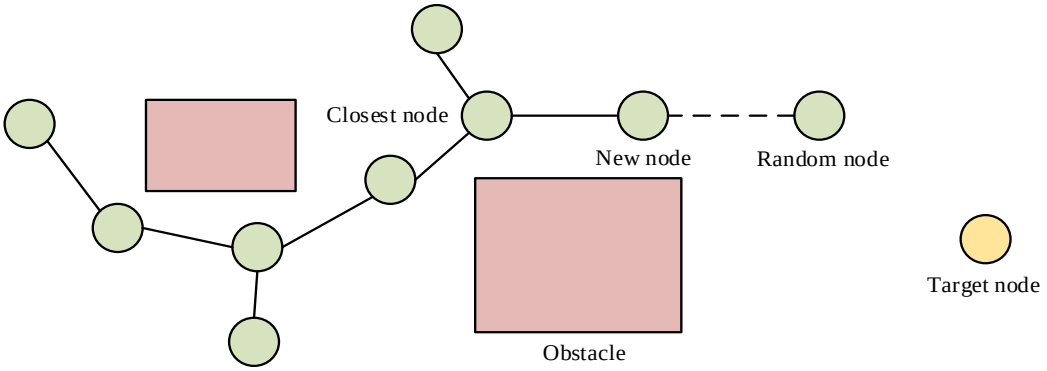


Fig. 4. Search tree extension plot of the RRT algorithm.

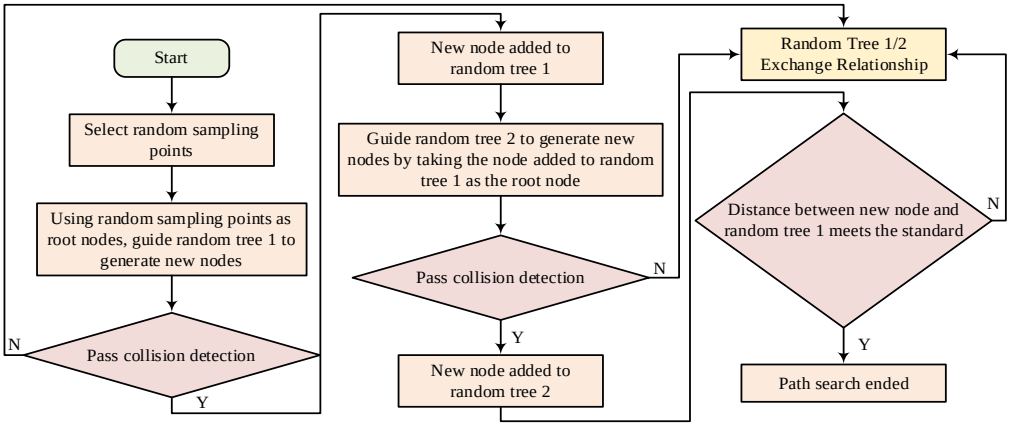


Fig. 5. The RRT-Connect path planning process.

From Fig. 5, firstly, a random sampling point is selected and used as the root node to guide random tree 1. A new node is generated. If the new node has not collided, it will be added to random tree 1. Otherwise, the node will be abandoned. After adding a new node to random tree 1, guide random tree 2 to generate a new node with that node as the root node. If the new node passes collision detection, it will be added to random tree 2. Otherwise, abandon the new node. After adding a new node to random tree 2, if the distance between the node and random tree 1 meets the standard, complete the path search. Otherwise, the relationship between the two random trees will be swapped. The improved RRT-Connect algorithm can obtain discrete path nodes. However, in the actual motion of the inspection robot, the trajectory of the push rod length needs to be planned to control the motion of the inspection robot. The trajectory of the push rod planned by the quintic polynomial interpolation method is shown in Eq. (13).

$$d(t) = m_0 + m_1 t + m_2 t^2 + m_3 t^3 + m_4 t^4 + m_5 t^5 \tag{13}$$

In Eq. (13), t represents the movement time of the push rod. m_0 - m_5 represent polynomial coefficients. The initial time is 0. The boundary condition is displayed in Eq. (14).

$$\begin{cases} d_1 = m_0 \\ v_1 = m_1 \\ a_1 = 2m_2 \\ d_2 = m_0 + m_1 t_2 + m_2 t_2^2 + m_3 t_2^3 + m_4 t_2^4 + m_5 t_2^5 \\ v_2 = m_1 + 2m_2 t_2 + 3m_3 t_2^2 + 4m_4 t_2^3 + 5m_5 t_2^4 \\ a_2 = 2m_2 + 6m_3 t_2 + 12m_4 t_2^2 + 20m_5 t_2^3 \end{cases} \tag{14}$$

In Eq. (14), d_1 and d_2 represent the initial length and termination length of the push rod, respectively. v_1 and v_2 represent the initial speed and end speed of the push rod. a_1 and a_2 are the initial acceleration and termination acceleration, respectively. t_1 and t_2 represent the initial and ending times, respectively. The coefficients of the five term equation can be obtained from the above equation, as shown in Eq. (15).

$$\begin{cases} m_0 = d_1 \\ m_1 = v_1 \\ m_2 = \dfrac{a_1}{2} \\ m_3 = \dfrac{20(d_2 - d_1) - (12v_1 + 8v_2)t_2 - (3a_1 - a_2)t_2^2}{2t_2^3} \\ m_4 = \dfrac{-30(d_2 - d_1) + (16v_1 + 14v_2)t_2 + (3a_1 - 2a_2)t_2^2}{2t_2^4} \\ m_5 = \dfrac{12(d_2 - d_1) - (6v_1 + 6v_2)t_2 - (a_1 - a_2)t_2^2}{2t_2^5} \end{cases} \tag{15}$$

Based on the quintic polynomial interpolation method, the speed, length, motion speed, and acceleration of the push rod are controlled, thereby achieving motion control of the inspection robot.

IV. Simulation Analysis of Obstacle Crossing Performance for TLIR

To test the obstacle crossing ability of the TLIR, simulation testing was performed. The obstacles selected in the experiment were the vibration damper and suspension clamp and insulator string. The lengths of the enclosure box for the vibration damper, suspension clamp, and insulator string were 50mm, 160mm, and 1150mm, respectively. The radii were 35mm, 45mm, and 100mm, respectively. The step size of the workspace was 30mm. The joint configurations of random tree root nodes for different obstacles were shown in Table I.

TABLE I. Joint Configurations of Random Root Nodes for Different Obstacles

Node	θ1/rad	d2/mm	θ3/rad	θ4/rad	θ5/rad	d6/mm	θ7/rad
Starting point of vibration damper	0.24	494.3	-0.24	0	-0.24	494.3	3.38
Midpoint of vibration damper	0.61	550.1	-0.61	0.44	-0.61	550.1	3.23
End of vibration damper	1.05	660.1	-1.05	0	-1.05	660.1	4.19
Starting point of insulator string	0.24	494.3	-0.24	0	-0.24	494.3	3.38
Midpoint of insulator string	0.56	586.0	-0.75	0.79	-0.56	577.0	2.71
End of insulator string	0.79	660.4	-0.79	0	-0.79	660.4	3.93

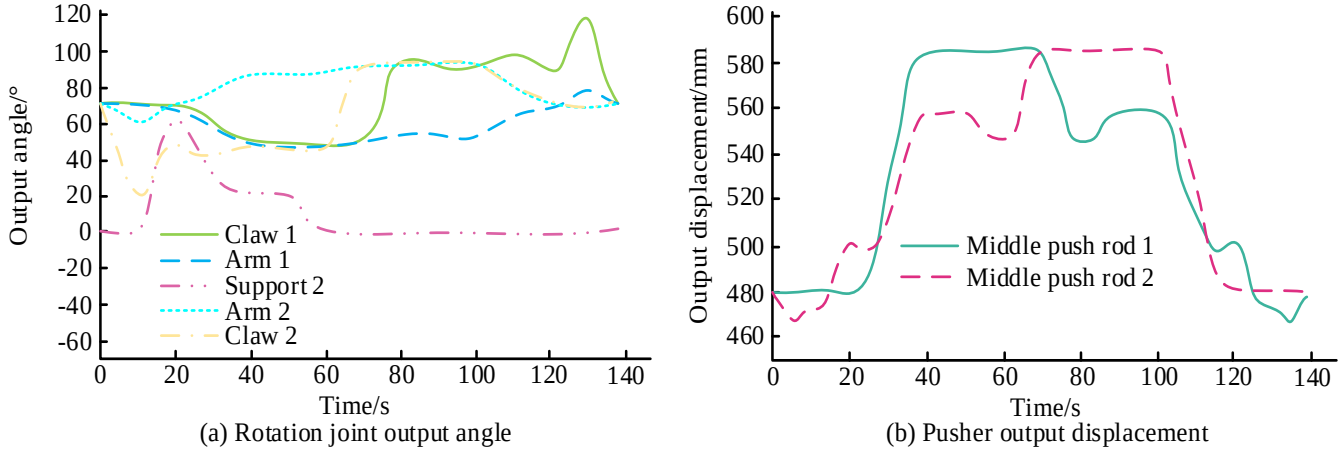
From Table I, when the inspection robot crossing different obstacles, the root node joint configuration of its random tree was different. When crossing the suspension clamp and insulator string, the output curve of the inspection robot is shown in Fig. 6.

In Fig. 6(a), when the crossing the suspension clamp and insulator string, the angle output range of claw 1 and arm 1 was (50°~120°) and (50°, 80°). The angle output ranges of the claw 2, bracket 2, and arm 2 were (20°, 90°), (0°, 60°), and (-60°, 0°). From Fig. 6(b), during the obstacle crossing process, the length output range of the push rods in numbers 1 and 2 was (468585)mm. According to Fig. 6, the end trajectory in the rear arm obstacle crossing is based on the front claw center as the reference point for the end base coordinate, and its path is the same as the front claw obstacle crossing trajectory. This proves that the symmetrical structure of the robot can simplify the obstacle crossing process by making the process of crossing obstacles with the backward paw the reverse movement of the front paw when crossing simple obstacles. The displacement of

each joint and push rod when the inspection robot crossed the suspension clamp and insulator string is shown in Fig. 7.

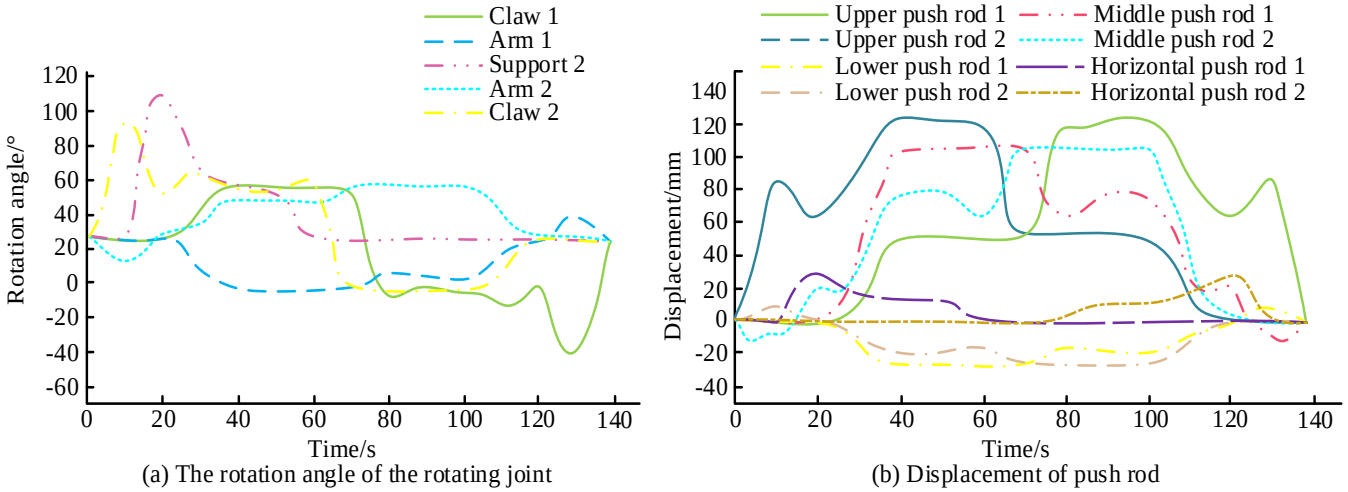
In Fig. 7(a), when crossing the suspension clamp and insulator string, the rotation ranges of claw 1 and arm 1 were (-50°~20°) and (-20°, 10°). The rotation ranges of claw 2, bracket 2, and arm 2 were (-20°, 50°), (0°, 60°), and (-10°, 20°). From Fig. 7(b), during the obstacle crossing process, the extension range of push rods 1 and 2 was (0, 125) mm. The extension range of the middle push rods 1 and 2 was (-15, 115) mm. The extension range of the lower push rods 1 and 2 was (-25, 0) mm. The extension range of the horizontal push rods 1 and 2 was (0, 30) mm. From Fig. 7, the motion process of the front claw under the reference mark when crossing obstacles is opposite to that of the rear claw when crossing obstacles under the reference mark. If the starting point of the obstacle crossing at a distance of 500mm from the hanging clamp on the wire is taken as the base coordinate, the simplified model of the rotation joint rotation angle and push rod displacement change process in the front claw obstacle crossing is mirrored to the rotation joint rotation angle and push rod displacement change process in the rear claw motion. From this, the robot's obstacle crossing movement is relatively stable. The output speed of each joint and push rod of the inspection robot is shown in Fig. 8.

From Fig. 8(a), when crossing the suspension clamp and insulator string, the rotational speeds of the claw 1 and arm 1 were (-7.5~7.5) deg/s and (-1.5, 1.5) deg/s. The rotational speeds of the claw 2, bracket 2, and arm 2 were (-7.5~7.5) deg/s, (-4, 11.5) deg/s, and (-1.5, 1.5) deg/s. From Fig. 8(b), during the obstacle crossing process, the telescopic speeds of push rods 1 and 2 were both (-13, 10) mm/s. The extension speeds of push rods 1 and 2 were both (-7.5, 8.5) mm/s. The telescopic speeds of the lower push rods 1 and 2 were both (-2.5, 2.5) mm/s. The telescopic speed of the horizontal push rods 1 and 2 was (-2.5, 6) mm/s. The speed change curve of each joint and push rod of the inspection robot was smooth, indicating that the acceleration value was not large. The motion of the inspection robot was relatively stable. When climbing over the shock absorber, the output curve of the inspection robot is shown in Fig. 9.



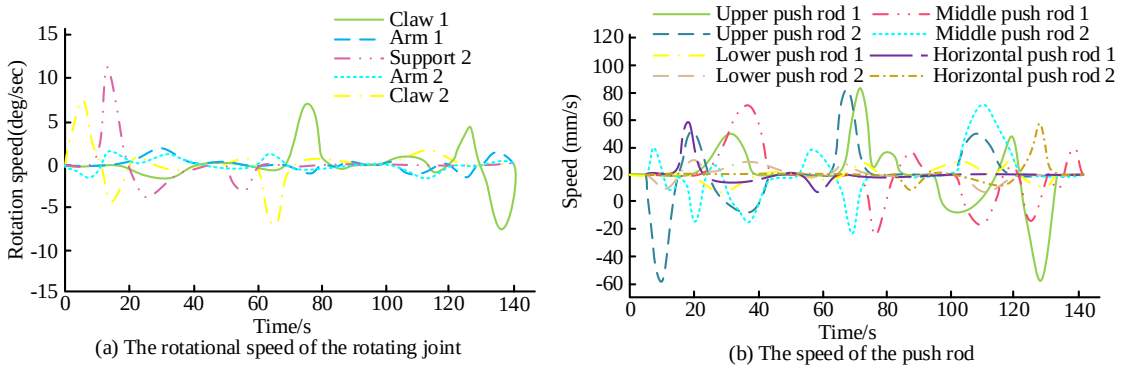
(a) Rotation joint output angle (b) Pusher output displacement

Fig. 6. Output curve of the inspection robot.



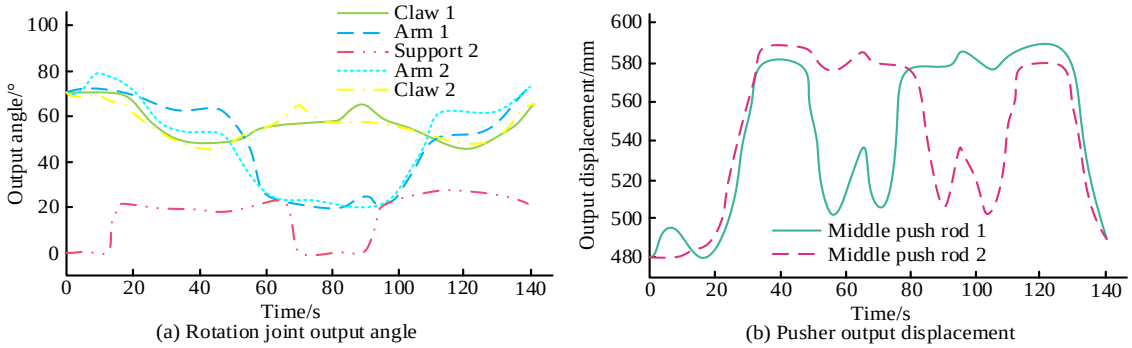
(a) The rotation angle of the rotating joint (b) Displacement of push rod

Fig. 7. Displacement of the joints and push rods of the inspection robot.



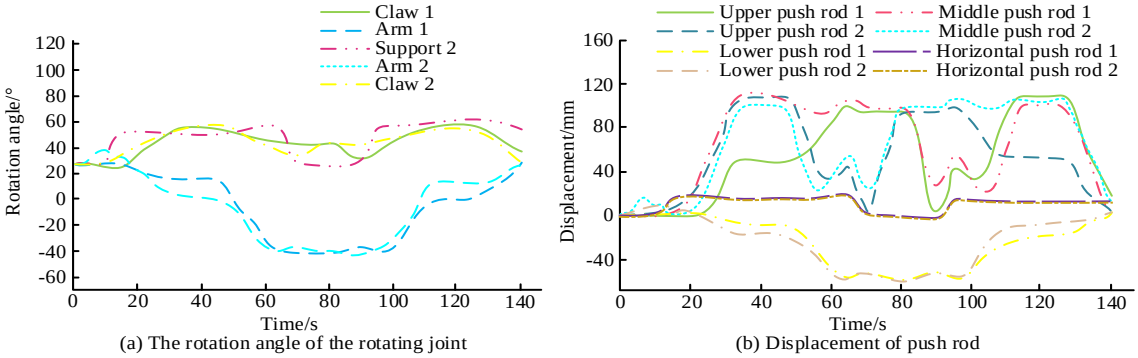
(a) The rotational speed of the rotating joint (b) The speed of the push rod

Fig. 8. Output speed of each joint and push rod of the inspection robot.



(a) Rotation joint output angle (b) Pusher output displacement

Fig. 9. Output curve of the inspection robot.



(a) The rotation angle of the rotating joint (b) Displacement of push rod

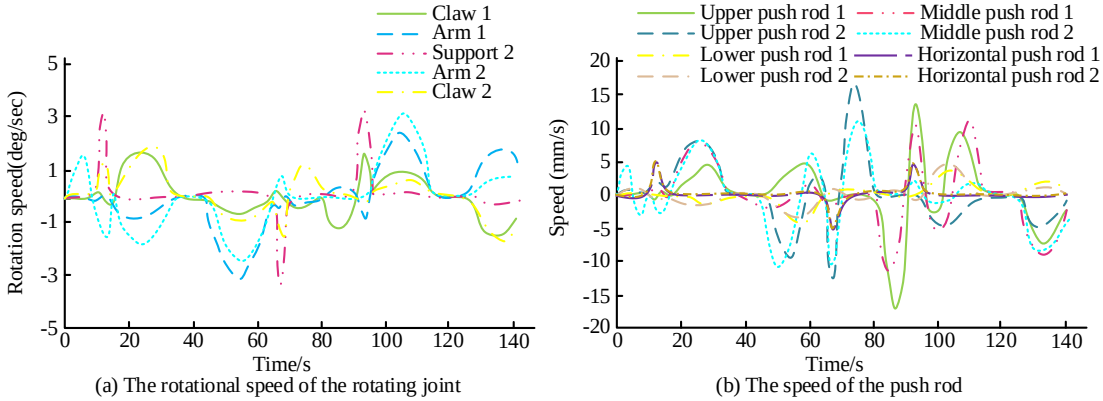
Fig. 10. Displacement of the joints and push rods of the inspection robot.

In Fig. 9(a), when overtaking the vibration damper, the angle output ranges of claw 1 and arm 1 were (50°~70°) and (20°, 80°). The angle output ranges of the claw 2, bracket 2, and arm 2 were (45°, 70°), (0°, 25°), and (0°, 25°). From Fig. 9(b), during the obstacle crossing process, the length output range of push rods 1 and 2 was (480, 590) mm. From Fig. 9, the obstacle crossing trajectory of the robot's front and rear claws over the shock absorber is basically consistent with the planned end trajectory. In order to save time, the trajectory of the rear claw passing through the end of the shock absorber does not need to be completely mirrored with the front claw, which can make it closer to the turning point of the planned path. When climbing over the vibration damper, the displacement of each joint and push rod is shown in Fig. 10.

From Fig. 10(a), the rotation ranges of claw 1 and arm 1 were (0°~22°) and (-50°, 10°) when overtaking the vibration hammer. The rotation ranges of the bracket 2, arm 2, and claw 2 were (0°, 25°), (-50°, 10°), and (0°, 22°). From Fig. 10(b), during the obstacle crossing process, the extension range of push rods 1 and 2 was (0, 100) mm. The extension range of the push rods 1 and 2 was (0, 110) mm. The extension ranges of the lower push rods 1 and 2 were (-60, 10) mm. The extension ranges of the horizontal push rods 1 and 2 were (0, 20) mm. When climbing over the vibration hammer, the output speed of each joint and push rod of the inspection robot is shown in Fig. 11.

From Fig. 11(a), the rotational speeds of the claw 1 and the arm 1 were (-1.5~1.5) deg/s and (-3, 2.5) deg/s when overtaking the damper. The rotational speeds of the bracket 2, arm 2, and claw 2 were (-3.5~3.5) deg/s, (-2.5, 3) deg/s, and (-27, 2) deg/s. From Fig. 11(b), during the obstacle crossing process, the extension speeds of push rods 1 and 2 were (-17, 13.5) mm/s and (-13.5, 17) mm/s, respectively. The extension ranges of push rods 1 and 2 were (-11, 11) mm/s. The extension speeds of

the lower push rods 1 and 2 were (-4.5, 3) mm/s and (-3, 4.5) mm/s, respectively. The stretching and retracting speeds of the horizontal push rods 1 and 2 were all (-4, 5) mm/s. When the inspection robot climbed over the vibration hammer, the speed change curve of each joint and push rod was smooth. It indicated that the motion of the inspection robot was relatively stable.



(a) The rotational speed of the rotating joint (b) The speed of the push rod

Fig. 11. Output speed of each joint and push rod of the inspection robot.

V. DISCUSSION

As an indispensable part of social production activities and people's lives, electricity plays an irreplaceable role in the development of the national economy. Overhead transmission lines are the hub responsible for transportation in the entire power system. It is necessary to conduct regular inspections and maintenance of high-voltage overhead transmission lines that are exposed to harsh natural environments all year round to ensure the safe and stable operation of the power transportation network. Compared with manual inspection, helicopter inspection, and drone inspection, online inspection robots have advantages such as low cost, high quality, and long operating time. To this end, the study conducted kinematic analysis on the inspection robot to understand its motion patterns during obstacle crossing. An obstacle crossing path planning algorithm based on an improved RRT-Connect algorithm was proposed. The experimental results showed that when the inspection robot crossed the insulator string of the suspension clamp, the angle output ranges of the first claw and the first arm were (50°~120°) and (50°, 80°). The angle output ranges of claw 2, bracket 2, and arm 2 were (20°, 90°), (0°, 60°), and (-60°, 0°). The corresponding rotational speeds were (-7.5~7.5) deg/s, (-1.5, 1.5) deg/s, (-7.5~7.5) deg/s, (-4, 11.5) deg/s, and (-1.5, 1.5) deg/s. The extension and contraction ranges of the upper/middle/lower/horizontal push rods were (0125) mm, (-15, 115) mm, (-25, 0) mm, and (0, 30) mm, respectively, with corresponding extension and contraction speeds of (-13, 10) mm/s, (-7.5, 8.5) mm/s, (-2.5, 2.5) mm/s, and (-2.5, 6) mm/s. According to the experimental results, the proposed inspection robot has stable changes in the angle (displacement) and velocity curves of its joints and push rods when crossing obstacles, without any abrupt changes. The same conclusion can also be observed from the graph of velocity and acceleration. At the same time, the velocity curves of push rods in each joint of the robot are smooth, and the acceleration values are not large. Therefore, the robot's obstacle crossing motion is relatively stable and has good kinematic performance. In addition, the output changes of each push rod obtained in the end path planning of the line patrol robot can enable the robot to complete the planned action along the predetermined trajectory at the obstacle crossing end.

VI. CONCLUSION

As the key to power transmission, transmission lines need to be regularly inspected to ensure safety and stability. The inspection robot has low cost and long running time in the inspection of transmission lines. It can promptly troubleshoot faults on transmission lines. To ensure that the inspection robot can achieve collision free motion in the inspection project, the kinematic analysis is conducted on the inspection robot to analyze the motion laws during obstacle crossing. An obstacle crossing path planning algorithm based on the improved RRT-Connect algorithm is proposed. In order to verify the effectiveness of the obstacle crossing method proposed in the study, simulations were conducted based on the obstacle crossing path of the robot's end effector. Real-time rotation angles, output displacement, and velocity curves of each joint and push rod during different obstacle crossing processes were obtained. Kinematic equations were used to further represent the variation curves of each joint. The output changes of each push rod obtained in the end path planning of the line inspection robot can enable the robot to complete the planned action along the predetermined trajectory at the obstacle crossing end, and ensure the stability of the inspection robot during obstacle crossing. Although the obstacle crossing method proposed in the study can effectively ensure the motion stability of the inspection robot, there are still many redundant trajectories in its workspace. Therefore, future work will focus on how to eliminate redundant trajectories.

REFERENCES

[1] Luo X J, Zhang LL. Adaptive Morphological Filter for Extracting Features of Fault on Transmission Line. IEEJ Transactions on Electrical and Electronic Engineering, 2022, 17(9):1364-1366.

[2] Viswavandya M, Patel S, Sahoo K. Analysis and Comparison of Machine Learning Approaches for Transmission Line Fault Prediction in Power Systems. Journal of Research in Engineering and Applied Sciences, 2021, 6(1):24-31.

[3] Nie Y, Liu Y, Lu D, & Wang B. An improved natural frequency based transmission line fault location method with full utilization of frequency spectrum information. IET Generation, Transmission & Distribution, 2021, 15(19):2787-2803.

[4] Hernando M, Brunete A, Gambao E. ROMERIN: A Modular Climber Robot for Infrastructure Inspection. IFAC-PapersOnLine, 2019, 52(15):424-429.

[5] Wang Y, Yuan C, Zhai Y. Mechanism Design and Analysis of a New Overhead Transmission Line Inspection Robot. Journal of Northwestern Polytechnical University, 2020, 38(5):1105-1111.

[6] Wei J, Hua Z D, Shuangbao M, Gaocheng Y, & Wei C. Dynamic walking characteristics and control of four-wheel mobile robot on ultra-high voltage multi-split transmission line. Transactions of the Institute of Measurement and Control, 2022, 44(6):1309-1322.

[7] Huang X, Wu Y, Zhang Y, & Zhang H. A Method of Transmission Conductor Loosened Detect Based on Image Sensors. IEEE Transactions on Instrumentation and Measurement, 2020,69P(11):8783-8796.

[8] Qingkai Z, Qingwu L, Chang X, Qiuyu L, & Yaqin Z. Class-aware edge-assisted lightweight semantic segmentation network for power transmission line inspection. Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies, 2023, 53(6):6826-6843.

[9] Xiong S, Liu Y, Yan Y, Pei L, Xu P, & Fu X, et al. Object recognition for power equipment via human ‐‐ evel concept learning. IET Generation, Transmission & Distribution, 2021, 15(10):1578-1587.

[10] Hu B, Cao Z, Zhou M. An Efficient RRT-based Framework for Planning Short and Smooth Wheeled Robot Motion under Kinodynamic Constraints. IEEE Transactions on Industrial Electronics, 2020, 68(4):3292-3302.

[11] Sun Y, Zhang C, Sun P, & Liu C. Safe and Smooth Motion Planning for Mecanum Wheeled Robot Using Improved RRT and Cubic Spline. Arabian Journal for Science and Engineering. Section A, Sciences, 2020, 45(4):3075-3090.

[12] Jhang J H, Lian F L, Hao Y H. Human-like motion planning for autonomous parking based on revised bidirectional rapidly-exploring random tree with Reeds-Shepp curve. Asian Journal of Control, 2021, 23(3):1146-1160.

[13] Zammit C, Kampen E J V. Real-time 3D UAV Path Planning in Dynamic Environments with Uncertainty. Unmanned Systems, 2023, 11(3):203-219.

[14] Wang J, Hirota K, Wu X,Dai Y, & Jia Z. Hybrid Bidirectional Rapidly Exploring Random Tree Path Planning Algorithm with Reinforcement Learning. Journal of Advanced Computatioanl Intelligence and Intelligent Informatics, 2021, 25(148):121-129.

[15] Choudhuri S, Adeniye S, Sen A. Distribution Alignment Using Complement Entropy Objective and Adaptive Consensus-Based Label Refinement for Partial Domain Adaptation, Artificial Intelligence and Applications. 2023, 1(1): 43-51.

[16] Belaid A, Mendil B, Djenadi A. Narrow passage RRT*: a new variant of RRT*. International journal of computational vision and robotics, 2022, 12(1):85-100.

[17] Huang G, Ma Q. Research on Path Planning Algorithm of Autonomous Vehicles Based on Improved RRT Algorithm. International journal of intelligent transportation systems research, 2022, 20(1):170-180.

[18] Wang X, Xia Z, Zhou X, Wei J, Gu X, & Yan H. Collision-Free Path Planning for Arc Welding Robot Based on Ida-De Algorithm. International Journal of Robotics & Automation, 2022, 37(6):476-485.

[19] Babel L. Online flight path planning with flight time constraints for fixed-wing UAVs in dynamic environments. International journal of intelligent unmanned systems, 2022, 10(4):130-157.

[20] Christensen R S, Droge G, Leishman R C. Closed-Loop Linear Covariance Framework for Path Planning in Static Uncertain Obstacle Fields. Journal of Guidance, Control, and Dynamics: A Publication of the American Institute of Aeronautics and Astronautics Devoted to the Technology of Dynamics and Control, 2022, 45(4):669-683.

Precision Machining of Hard-to-Cut Materials: Current Status and Future Directions

Tengjiao CUI
Shijiazhuang Institute of Technology, Shijiazhuang 050228, China
Shijiazhuang Tiedao University SiFang College, Shijiazhuang 051132, China

***Abstract*—Machining difficult materials like superalloys, ceramics, and composites is fundamental in industries where performance is paramount, such as the auto industry, aerospace, and medicine. These materials with relatively high strength, hardness, and high-temperature capabilities pose difficulties in machining, thus calling for improved precision machining technologies. This survey paper presents a detailed review of the current state of the art of precision machining of these difficult materials, along with advances observed in tools for cutting, machining techniques, and new technologies. They range from carbide, ceramics, super hard tools, and geometry of tools, and this topic also deals with tool coatings. The article also discusses specifics of the traditional and nontraditional machining processes: turning, milling, electrical discharge, and laser machining, as well as the relations between additive and hybrid manufacturing. The importance of new technologies or digital and intelligent manufacturing systems in enhancing the accuracy and productivity of machining is also illustrated. Furthermore, the paper also provides information on how digital and intelligent manufacturing technologies can enhance machining efficiency and accuracy. Moreover, future research will aim to minimize tool wear, enhance surface finish and integrity, and environmentally conscious machining. The paper concludes with a hopeful note on the potential of future research to revolutionize the precision machining industry, offering high performance and reliability in critical applications while maintaining a focus on sustainability.**

Keywords*—*Precision machining; hard-to-cut materials; cutting tools; machining processes; emerging technologies

I. Introduction

A. Context

Nowadays, many industrial sectors require construction supplies characterized by superior mechanical and functional features [1]. Construction materials exhibit exceptional hardness, abrasion resistance, temperature-resistant strength, enhanced thermal conductivity, and oxidation and corrosion resistance and are indispensable in a variety of high-demand industries, including aerospace [2], biomedical [3], electronics, and automotive [4]. Despite these unique features, these materials are regarded as hard to cut because of their poor machinability [5].

Recent advancements in cutting tools, including carbide, ceramics, and super-hard tools, have significantly improved the machinability of hard-to-cut materials [6]. Innovative tool geometries and coatings have further enhanced machining performance. Moreover, conventional and non-conventional machining methods such as turning, milling, Electrical Discharge Machining (EDM), and laser machining are gaining attention for their role in addressing these challenges [7]. Integrating hybrid machining techniques, including additive manufacturing, has opened new avenues for achieving greater precision and customization in the machining process [8].

B. Challenges

There are several major problems associated with machining hard-to-cut components, including the high levels of machining forces [9], the high levels of vibrations encountered in machining systems [10], the concentration of heat, the rapid increase in machining temperature [11], rapid tool wear, and catastrophic tool failure [12], as well as frequent instability loss and significant deterioration of surface finishes [13].

Maintaining surface integrity while improving efficiency is a critical challenge that hinders the machining of hard-to-cut materials [14]. Moreover, with increasing demands for sustainable practices, addressing environmental concerns in machining processes remains largely underdeveloped [15]. Sustainable machining, aimed at reducing energy consumption, tool wear, and waste generation, is becoming increasingly essential, but solutions are still in the early stages of implementation [16].

C. Contribution

The primary aim of this paper is to deliver a thorough overview of current advances in precision machining of hard-to-cut steels, focusing on the challenges and innovations in cutting tools, machining processes, and new technologies. By analyzing recent advances in material compositions, coatings, and geometries, as well as both conventional and non-conventional machining techniques, this article aims to provide comprehensive insights into state-of-the-art methods and practices. In addition, an attempt is made to identify future research directions to address existing challenges such as tool wear, surface integrity, and sustainability. Ultimately, the goal is to support continuous improvement in machining efficiency, accuracy, and sustainability in industries that rely on these demanding materials.

The novelty of this study lies in its exploration of integrating digital and intelligent manufacturing technologies to improve machining efficiency and accuracy. By reviewing these

advancements and suggesting future research avenues, the paper contributes to ongoing efforts to improve precision machining for high-demand applications across various industries.

II. Characteristics of Hard-to-Cut Materials

Materials with challenging machining characteristics possess exceptional mechanical and thermal properties, which make them indispensable for high-performance applications but also pose significant challenges during machining [17, 18]. As illustrated in Fig. 1, materials like ceramics, nickel-based superalloys, titanium alloys, and composites exhibit varying degrees of hardness, strength, thermal resistance, and other critical properties that directly impact their machinability. These inherent properties lead to issues like rapid tool wear, high cutting forces, and poor machinability, which compares key challenges across materials. Consequently, machining these materials requires specialized cutting tools and advanced techniques tailored to their specific characteristics [19]. Understanding these unique properties, as detailed in Table I, is essential for developing effective machining strategies that ensure precision, efficiency, and the desired surface quality of manufactured components.

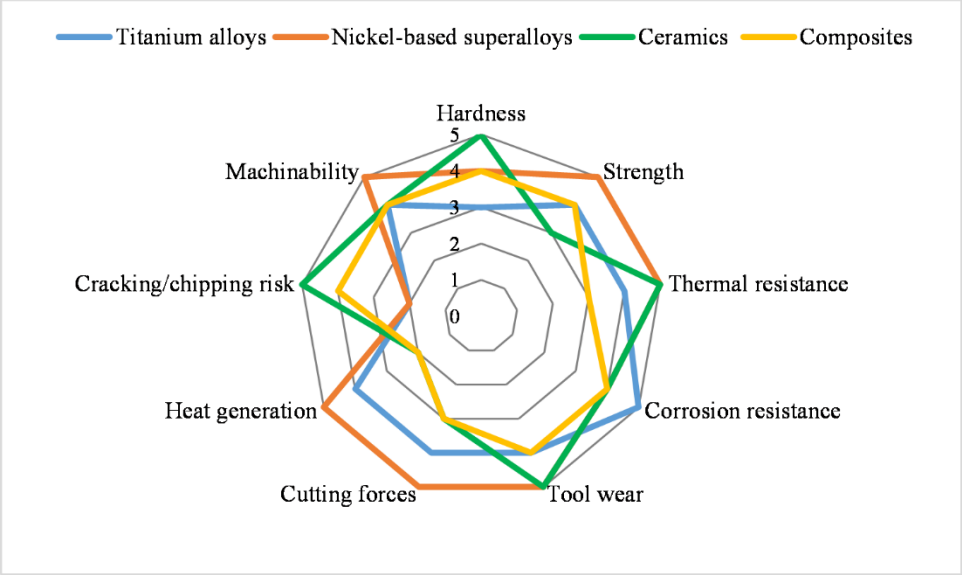


Fig. 1. Key properties and machining challenges of hard-to-cut components.

TABLE I. Comparative Characteristics and Machining Challenges of Hard-to-Cut Materials

Materials	Properties				Challenges					Relevant industries			
	High hardness	High strength	Thermal resistance	Corrosion resistance	Rapid tool wear	High cutting forces	Heat generation	Risk of cracking/chipping	Poor machinability	Aerospace	Automotive	Biomedical	Power Generation
Titanium alloys	Medium	High	High	Very high	High	High	High	Low	High	✓	✓	✓	
Nickel-based superalloys	High	Very high	Very high	High	High	Very high	Very high	Low	Very high	✓	✓		✓
Ceramics	Very high	Medium	Very high	High	Very high	Medium	Low	Very high	High	✓	✓	✓	
Composites	High	High	Medium	High	High	Medium	Low	High	High	✓	✓	✓	

A. Titanium Alloys

Titanium alloys, known for their outstanding biocompatibility, corrosion resistance, and strength-to-weight percentage, have become essential in aerospace, biomedical, and chemical processing [20]. Despite these advantages, machining titanium alloys presents significant challenges with cutting tools, given their strong chemical reactivity and limited thermal conductivity. These properties result in heat concentration in cutting areas, causing rapid tool wear and possible workpiece damage.

In addition, the high strength of titanium alloys increases cutting forces, further exacerbating tool wear and shortening tool life [21]. To mitigate these problems, various strategies are used, such as using advanced tool materials such as cemented carbide and ceramics, high-pressure cooling systems, and

optimizing cutting parameters. Innovations in tool coatings such as Titanium Aluminum Nitride (TiAlN) have also shown promise in improving tool performance and longevity. Understanding the machinability characteristics of titanium alloys is critical to developing efficient and precise machining procedures that satisfy the rigorous demands of high-performance applications [22].

B. Nickel-based Superalloys

Nickel-based superalloys like Hastelloy and Inconel are important materials for high-temperature and high-voltage uses, especially in the aerospace and power generation fields [23]. These alloys are valued for their superior mechanical characteristics under high temperatures, including their strength, stability, and resistance to corrosion. However, due to several factors, such materials are notoriously hard to machine. Nickel-based superalloys' thermal strength and work-hardening properties result in significant cutting forces and rapid tool wear.

Moreover, their poor thermal conductivity results in heat accumulating in cutting zones, increasing tool wear and potentially compromising the workpiece's surface integrity. Advanced machining techniques, such as super-hard cutting tools (e.g., cubic boron nitride), optimized cutting parameters, and high-pressure coolant systems, are essential to addressing these challenges. Coatings such as TiAlN and AlCrN on carbide tools can improve tool life and performance [24]. Understanding machinability issues and implementing effective machining strategies are critical to successfully manufacturing nickel-based superalloy components, ensuring reliability and performance in demanding applications.

C. Ceramics

Ceramics, such as cutting tools, medical devices, and aerospace components, are commonly employed in products with high hardness, wear resistance, and thermal stability. Their exceptional properties are ideal for high-stress, high-temperature environments [25]. However, ceramics are also extremely brittle and have low fracture toughness, which makes them difficult to process. The brittleness of ceramic causes cracking and chipping to occur during processing, which can significantly affect the final product's surface quality and dimensional accuracy.

Conventional machining methods often produce inadequate material removal and excessive tool wear coefficients. Advanced techniques like Laser-Assisted Machining (LAM), EDM, and Ultrasonic Machining (USM) are increasingly adopted to address these challenges. These methods help reduce cutting forces, increase material removal rates, and improve surface finish. Additionally, innovations in diamond-coated tools and fine-grain ceramic tool materials have shown promise in improving the machinability of ceramics. Understanding ceramics' unique properties and machining challenges is critical to developing effective machining strategies that ensure ceramic components' high precision and reliability.

D. Composites

Composites like Glass Fiber-Reinforced Polymers (GRP) and Carbon Fiber-Reinforced Polymers (CFRP) offer many advantages in the automotive, aerospace, and sports equipment fields for their excellent strength-to-weight balance, corrosion resistance, and adaptable properties [26]. Despite these advantages, machining composite materials presents significant challenges due to their heterogeneous and anisotropic nature. The different material properties within the composite structure lead to different wear on the cutting tools and cause problems, including matrix cracking, fiber pull-out, and delamination.

Conventional machining techniques often struggle to maintain the integrity of the fibers and the matrix, resulting in less accurate dimensional measurements and inferior surface finishes [27]. To mitigate these problems, advanced machining strategies like Abrasive Water Jet Machining (AWJM), laser machining, and ultrasonic-assisted machining are used. These methods provide greater control over the machining process, reducing damage and improving surface quality. In addition, developing special cutting tools with optimized geometries and coatings has improved the machinability of composite materials. Understanding the specific machining challenges and developing tailored strategies are critical to effectively and precisely manufacturing composite components, ensuring their performance and longevity in demanding applications.

III. Advances in Cutting Tools

Cutting tools have made significant progress, driven by the need to boost the machinability of hard-to-cut components. Modern cutting tools are designed to withstand extreme conditions when machining superalloys, ceramics, and composites. As detailed in Table II, tool materials, geometries, and coatings innovations have greatly improved tool efficiency, extended tool lifespan, and machining efficiency. The application of super-hard materials, including Cubic Boron Nitride (CBN) and Polycrystalline Diamond (PCD), as well as advanced coatings like Diamond-Like Carbon (DLC) and TiAlN, has resulted in better wear resistance and heat dissipation. Additionally, optimized tool geometries, including micro-textured surfaces and variable helix designs, have contributed to reduced cutting forces and improved chip evacuation. These advances are critical to overcoming hard-to-cut material challenges and enabling high-precision, high-efficiency machining processes.

A. Tool Materials

Carbide tools, ceramic tools, PCD, and CBN are among the most advanced materials used in cutting tools, thanks to their exceptional hardness, thermal stability, and wear resistance. As detailed in Fig. 2, these materials exhibit distinct features that make them useful for different machining applications. Understanding their comparative performance is essential for selecting the appropriate tool for specific machining tasks.

1) Carbide tools: Carbide tools are commonly used when cutting hard-to-cut materials for their extreme hardness, toughness, and wear resistance. Composed primarily of tungsten carbide particles bonded to a metallic cobalt matrix, these tools are designed to keep their cutting-edge and functional integrity even in extreme situations. Carbide tools are particularly effective in high-speed machining operations where they can withstand the heat generated without losing their hardness.

TABLE II. ADVANCES IN CUTTING TOOLS

Aspect	Description	Examples	Benefits
Carbide tools	Commonly used for their hardness, toughness, and wear resistance.	Coatings: TiAlN, TiCN, DLC. Improvements: Ultra-fine carbides, special bonding phases.	High-speed machining capabilities, reduced friction, improved thermal stability, and extended tool life.
Ceramic tools	Known for superior hardness, high-temperature stability, and resistance to wear and chemical erosion.	Materials: Si3N4, Al2O3, SiC. Innovations: Whisker-reinforced ceramics, fine-grained ceramics.	Maintains hardness at high temperatures, is suitable for high-speed machining, reduces the likelihood of thermal damage, and is excellent for interrupted cutting.
PCD tools	Polycrystalline diamond tools are ideal for non-ferrous metals, composites, and abrasive materials.	High hardness and thermal conductivity are excellent for surface finish and unsuitable for ferrous materials.	Longer tool life, higher cutting speeds, improved productivity, minimal delamination, and fiber pull-out in composites.
CBN tools	Cubic boron nitride tools, designed for ferrous materials, have high thermal stability and wear resistance.	Effective for hardened steels, cast iron, and superalloys.	Higher cutting speeds, reduced cycle times, improved efficiency, excellent precision, and tool longevity in high-stress conditions.
Optimized rake angles	Improves chip formation, reduces cutting forces, and minimizes tool wear.	Positive rake angles for better shearing action.	Smoother cuts, better surface finishes, improved tool life, and reduced machining forces.
Variable helix angles	Reduces vibration and chatter during machining.	Varying helix angle along the cutting edge.	A more stable cutting process, better material removal rates, tool life, and surface quality are needed.
Micro-textured surfaces	Reduces friction and improves lubrication.	Microscopic patterns or textures on cutting surfaces.	Lower heat generation, higher machining efficiency, less tool wear, and better surface finishes.
Specialized edge preparations	Enhances durability and performance of cutting tools.	Techniques: Honing, chamfering.	Increased tool strength, resistance to chipping, reduced crack initiation, and improved reliability and performance for brittle materials.
Customized geometries	Tailored to specific applications for optimal performance.	Features: Complex groove designs, variable pitch, advanced chip breakers.	Enhanced performance, reduced tool wear, improved machining efficiency, and better suitability for specific material and process requirements.

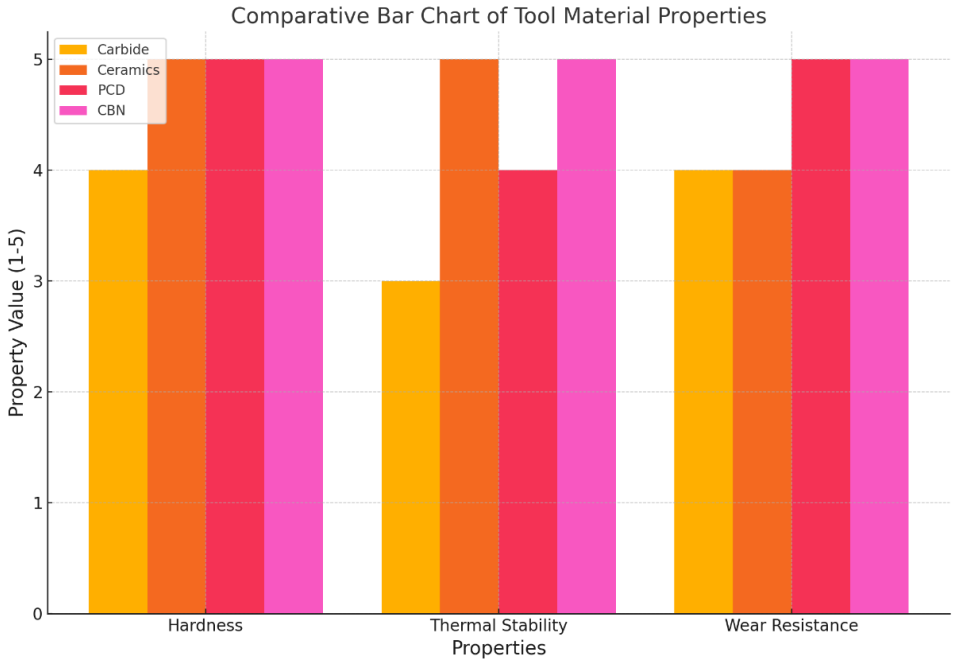


Fig. 2. Key properties of cutting tool materials.

The performance of carbide tools can be significantly enhanced by applying various coatings, improving wear resistance, thermal stability, and reducing friction. As illustrated in Fig. 3, different coatings like TiAlN, Titanium Carbonitride (TiCN), DLC, and AlCrN offer varying performance levels across these key metrics. Recognizing these variations is essential for choosing the most effective coating suited to the machining requirements.

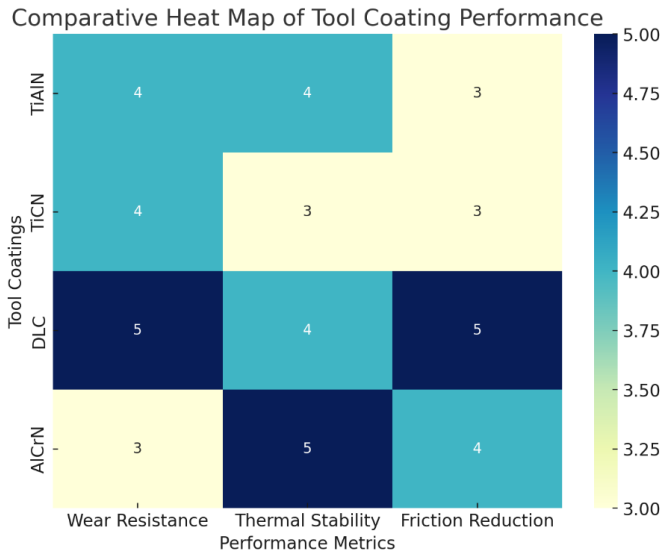


Fig. 3. Comparative heat map of tool coating performance.

Additionally, innovations in the microstructure and composition of cemented carbide tools have improved toughness and thermal cracking resistance. This includes the use of ultra-fine carbides that provide a balance between hardness and toughness and the incorporation of particular bonding phases to improve the overall performance of the tool.

2) Ceramic tools: Ceramic tools are highly effective at machining hard-to-cut items owing to their superior hardness, high-temperature stability, and wear and chemical erosion resistance. These tools are typically made of materials such as Silicon Carbide (SiC), Aluminum Oxide (Al2O3), and Silicon Nitride (Si3N4), which provide excellent mechanical properties even at elevated temperatures. This makes ceramic tools particularly suitable for high-speed machining applications where traditional tools may fail.

Ceramic tools can maintain hardness at temperatures that would soften other tool materials. This thermal stability enables higher cutting speeds and feeds, improving productivity while minimizing the thermal degradation of the workpiece. For example, silicon nitride-based ceramics are renowned for their toughness and resistance to thermal shock, enabling interrupted cutting operations.

Ceramic tools also benefit from advances in materials science, such as the development of whisker-reinforced ceramics. These materials contain silicon carbide whiskers in an aluminum oxide matrix, significantly improving fracture toughness and resistance to thermal shock. This makes them more robust under demanding machining conditions and extends their applicability to a wider range of hard-to-cut components.

However, ceramic tools' brittleness remains a challenge, as they are more prone to chipping and breaking under mechanical stress. To mitigate this, manufacturers have developed advanced tool geometries and surface treatments to improve the durability and performance of ceramic tools. For example, fine-grained ceramics and specially developed edge preparations help reduce the susceptibility to chipping.

3) Polycrystalline diamond and cubic boron nitride tools: PCD and CBN tools represent the pinnacle of cutting materials, distinguished by their exceptional hardness and wear resistance. These super-hard materials are specifically designed for the most demanding machining tasks with difficult-to-cut materials. PCD tools consist of fine diamond particles sintered in a metallic matrix under intense pressure and temperature. These tools have excellent thermal and hardness conductivity, providing excellent performance on non-ferrous metals, composites, and abrasive materials. The superior wear resistance of PCD tools enables longer tool life and higher cutting speeds, which translates into improved productivity and cost efficiency.

PCD tools are particularly effective in applications that require high surface quality, such as machining CFRP and aluminum alloys. Their ability to maintain sharp cutting edges ensures minimal delamination and fiber pull-out in composites while providing excellent surface finishes on metals. However, PCD tools are unsuitable for machining ferrous materials because chemical interactions at high temperatures lead to rapid tool wear. CBN tools are second only to diamonds in terms of hardness and are specifically designed for machining ferrous materials. CBN is synthesized by bonding cubic boron nitride grains with a ceramic or metallic binder. This creates a tool that can withstand extreme temperatures and maintain its cutting edge even under high-stress conditions.

CBN tools are particularly effective when machining hardened steels, cast iron, and superalloys. Their high thermal durability and abrasion resistance make them suitable for high-speed machining and finishing operations where precision and tool longevity are critical. CBN tools can operate at significantly higher cutting speeds than traditional carbide tools, reducing cycle times and improving overall efficiency. The development of advanced CBN grades and coatings has further enhanced their performance. These innovations have improved toughness and resistance to edge chipping, making CBN tools more versatile for a broader range of applications. In addition, new manufacturing techniques have enabled the production of CBN tools with complex geometries and expanded their applicability in precision machining.

B. Tool Geometries

Geometry affects the performance of cutting tools, particularly when dealing with hard-to-cut components. Advances in tool geometries have significantly improved cutting tool efficiency, precision, and life. Critical innovations in tool geometries include optimized rake angles, variable helix angles, and micro-textured surfaces designed to address specific machining challenges associated with difficult-to-machine materials. The rake angle of a cutting tool influences

chip formation, cutting forces, and heat generation. Optimized rake angles diminish cutting forces and minimize tool wear. Positive rake angles can improve the shearing action on difficult-to-cut metals, resulting in smoother cuts and better surface finishes. However, when choosing the rake angle, the sharpness must be balanced with the tool's strength to avoid premature failure.

Variable helix angles are designed to reduce vibration and chatter during machining, which is common when machining hard materials. By varying the helix angle along the cutting edge, these tools can disrupt the harmonic frequencies that cause chatter, resulting in a more stable cutting process. This improves surface quality, prolongs tool life, and results in higher material removal. Micro-textured tool surfaces are another significant advance in tool geometry. These tools have microscopic patterns or textures on their cutting surfaces to reduce friction and improve lubrication. The microtextures can serve as a reservoir for cutting fluids or can be used to break contact between the tool and the workpiece, thereby reducing heat generation and wear. This technology is particularly beneficial for machining difficult-to-machine materials where high temperatures and friction are a significant problem.

Edge preparation techniques such as honing and chamfering improve the durability and performance of cutting tools. For example, ground edges can improve tool strength and resistance to chipping, while beveled edges help reduce the formation of cracks in brittle tool materials such as ceramics and CBN. These unique edge treatments are tailored to the specific requirements of machining difficult-to-machine materials and ensure excellent reliability and performance. In addition to general advances, there is a growing trend toward customizing tool geometries for particular applications. This involves developing tools with unique features, such as complex groove design, variable pitch, and advanced chip breakers tailored to the particular requirements of the machining material and the machining process. Custom geometries can significantly increase performance, reduce tool wear, and improve machining efficiency.

IV. Machining Processes

Machining processes for hard-to-cut materials are constantly evolving to meet the unique challenges of these materials. Table III details that traditional machining methods, such as turning and milling, have been optimized with advanced techniques and technologies to increase their effectiveness. Non-conventional machining processes, including EDM, Laser Beam Machining (LBM), and USM, offer alternative approaches to achieving high precision and surface quality in these difficult-to-machine materials. Integration of these methods and innovations in machining strategies and process controls is essential to improving productivity, reducing tool wear, and ensuring the integrity of end components. This section covers conventional and non-conventional machining processes and highlights their applications, advantages, and advances.

The effectiveness of different machining processes in handling hard-to-cut materials varies significantly based on factors such as precision, surface finish, material removal rate, and tool wear. As shown in Fig. 4, conventional processes like turning and milling and non-conventional methods like EDM, LBM, and USM exhibit different strengths across these key metrics. Understanding these differences is crucial for selecting the appropriate machining process based on the specific material and application requirements.

TABLE III. Machining Processes for Hard-To-Cut Materials

Process	Description	Advantages	Limitations	Key Applications
Turning	Rotates the workpiece against a stationary cutting tool to remove material.	High precision, well-established, suitable for a variety of materials	Rapid tool wear, high cutting forces, heat generation	Aerospace, automotive, and biomedical components
Milling	Uses a rotating cutting tool to remove material from a stationary workpiece, creating complex geometries.	Versatile, high precision, capable of intricate details	High cutting forces, tool wear, heat generation	Aerospace, automotive, mold making, complex part manufacturing
EDM	Uses electrical discharges to erode the material, which is ideal for complex shapes and hard materials.	Precise, capable of intricate shapes, no mechanical stress	Slow process, limited to conductive materials, electrode wear	Turbine blades, injection molds, precision surgical instruments
LBM	Uses a focused laser beam to remove material, suitable for precise and complex shapes.	High precision, non-contact process, minimal heat-affected zone	High initial investment, material-specific challenges, potential thermal effects	Aerospace components, medical devices, microelectronic components, automotive parts
USM	Uses high-frequency ultrasonic vibrations with abrasives to remove material, which is effective for hard and brittle materials.	Non-thermal process, minimal mechanical stress, high precision	Slow material removal rate, tool wear, complex setup	Hard ceramic components, composite materials, precision surgical instruments, microelectronic components

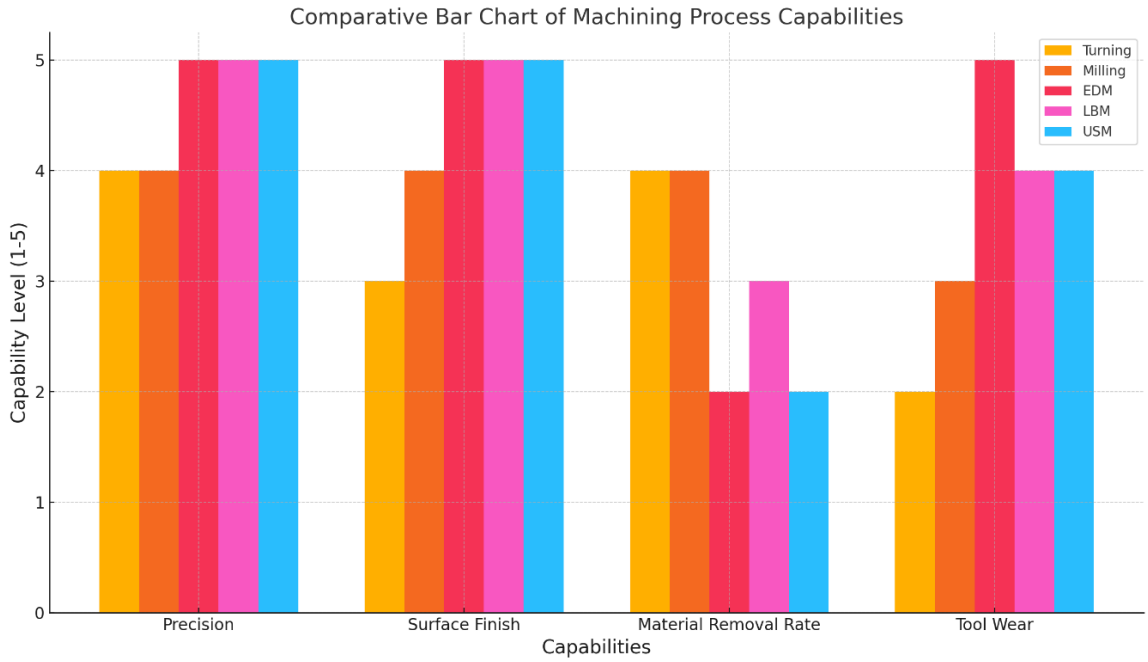


Fig. 4. Machining process capabilities.

A. *Conventional Machining*

1) Turning: Turning is a fundamental machining process often used to produce hard-to-cut materials. The workpiece is rotated against a stationary cutting tool, gradually removing material to achieve the desired shape and dimensions. While turning is a well-established process, machining difficult-to-machine materials such as titanium alloys, nickel-based superalloys, ceramics, and composites presents unique challenges that require specialized approaches and advanced technologies.

Turning hard-to-cut materials often leads to problems such as rapid tool wear, high cutting forces, and elevated temperatures in the cutting zone. These challenges can impact the surface finish, dimensional accuracy, and overall efficiency of the machining process. For example, the low thermal conductivity of titanium alloys causes heat to concentrate at the cutting edge, accelerating tool wear and potentially damaging the workpiece. Similarly, the work-hardening property of nickel-based superalloys increases cutting resistance, making it challenging to maintain tool sharpness and precision.

To overcome these challenges, using advanced tool materials and coatings is essential. Carbide tools with special coatings such as TiAlN and AlCrN offer improved wear resistance and thermal stability. Ceramic tools, known for their hardness and heat resistance, are also used for turning high-temperature materials. PCD and CBN tools are increasingly used due to their exceptional hardness and ability to maintain sharp edges, particularly in finishing operations.

Choosing the proper cutting parameters is crucial for successfully turning hard-to-cut materials. This includes optimizing cutting speed, feed rate, and depth of cut to balance tool life and machining efficiency. Lower cutting speeds are often used to reduce heat generation, while appropriate feed rates help control cutting forces and achieve a desired surface finish. High-pressure coolant systems effectively dissipate heat and lubricate the cutting zone, extending tool life and improving machining performance.

Innovative turning techniques such as multi-axis turning and high-speed turning have been developed to improve the performance and flexibility of the process. Multi-axis turning allows more complex geometries to be machined with greater precision and fewer setups. High-speed turning, made possible by advances in machine tool technology, helps reduce machining time and improve surface quality by operating at higher spindle speeds and feed rates. The integration of tool condition monitoring systems is becoming increasingly important. These systems use sensors and data analytics to monitor tool wear and predict failures, enabling timely tool changes and minimizing downtime. This real-time monitoring helps maintain consistent quality and improves overall process reliability.

2) Milling: Milling is a versatile machining process for shaping and contouring hard-to-cut materials. In milling, a rotating cutting tool removes material from a stationary workpiece, allowing the creation of complex geometries, slots, and intricate details. Despite its versatility, milling hard-to-cut materials such as titanium alloys, nickel-based superalloys, ceramics, and composites presents significant challenges due to their inherent properties.

Milling these materials requires high cutting forces, rapid tool wear, and significant heat generation. The brittleness of

ceramics and composites further complicates the process and increases the risk of chipping and delamination. Materials such as titanium and nickel-based superalloys have low thermal conductivity, resulting in heat concentration at the cutting edge, increasing tool wear, and compromising surface integrity.

To overcome these challenges, advanced tool materials such as carbide, ceramic, PCD and CBN are used. Carbide tools with special coatings such as TiAlN and DLC increase wear resistance and reduce friction, improving tool life and performance. Ceramic tools are ideal for high-speed milling operations due to their high hardness and thermal stability. PCD and CBN tools offer excellent edge retention and are particularly suitable for finishing operations in hard and abrasive materials.

The selection of suitable milling parameters is crucial for the efficient machining of hard-to-cut materials. This includes optimizing spindle speed, feed rate, and cut depth to balance tool life and machining efficiency. High-speed milling techniques coupled with advanced toolpath strategies help reduce cycle times and improve surface quality. Adaptive control systems and real-time monitoring further improve process stability and tool performance.

Innovative techniques such as trochoidal and high-feed milling were developed to meet the specific challenges of hard-to-cut materials. Trochoidal milling is a circular tool path that shortens the time the tool engages with the material, minimizing heat generation and tool wear. In high-feed milling, however, shallow cutting depths are used with high feed rates, which distributes the cutting forces more evenly and extends tool life.

Adequate coolant and lubrication strategies are essential when milling hard-to-cut materials to regulate heat and reduce friction. High-pressure coolant systems, cryogenic cooling, and Minimum Quantity Lubrication (MQL) techniques improve heat dissipation, tool life, and better surface finishes. These cooling methods are particularly beneficial for maintaining the tool's and the workpiece's structural integrity.

Implementing tool condition monitoring systems in milling operations enables real-time tool wear and performance assessment. These systems can use sensors and data analysis to predict failures and optimize tool change intervals, minimizing downtime and ensuring consistent machining quality. This proactive approach increases overall process reliability and efficiency.

B. Non-Conventional Machining

1) Electrical discharge machining: EDM is an unconventional process that has become an important method for machining hard-to-cut materials. EDM uses electrical discharges or sparks to remove material from the workpiece. This enables precise machining of complex geometries and hard materials that are difficult to machine using traditional methods. This process is particularly effective for hardened steels, superalloys, and conductive ceramics.

EDM works on the principle of spark erosion. The workpiece and electrode are immersed in a dielectric fluid, usually deionized water or oil. When a voltage is applied, a spark discharge occurs in the small gap between the electrode and the workpiece, generating intense heat (up to 12,000 °C) that melts and vaporizes a small portion of the workpiece material. The dielectric fluid cools and washes away the eroded particles, providing a clean machining environment.

There are two main types of EDM: wire EDM and die-sinking EDM. Wire EDM uses a thin, electrically conductive wire as an electrode to cut through the workpiece, making it ideal for producing intricate shapes and fine details. Die-sinking EDM, on the other hand, uses a pre-formed electrode to create cavities and complex contours in the workpiece, and it is commonly used in mold and die-making.

EDM offers several advantages when machining hard-to-cut materials. It can produce complicated shapes and high-precision components with excellent surface finish, regardless of the material's hardness. The absence of mechanical forces during machining eliminates problems related to tool wear and deformation common in traditional machining processes. EDM is also suitable for processing heat-treated materials without causing thermal deformation or internal stresses.

Despite its advantages, EDM has certain limitations. The process is relatively slow compared to traditional machining methods and is less suitable for mass production. In addition, EDM is limited to electrically conductive materials, which limits its scope. Electrode wear can also be a problem and requires frequent replacement and careful selection of electrode materials.

Recent advances in EDM technology focus on improving process efficiency, precision, and surface quality. Innovations such as high-speed EDM, multi-axis EDM, and adaptive control systems have improved the capabilities of EDM. High-speed EDM machines utilize advanced power and control systems to increase material removal rates and reduce machining time. Multi-axis EDM enables more complex geometries and greater accuracy by allowing simultaneous movements along multiple axes. Adaptive control systems optimize machining parameters in real time, improving process stability and tool life.

EDM is widely used in the aerospace, automotive, medical, and tool and mold industries. It is particularly valuable for producing components with complicated geometries, fine features, and high surface quality requirements. Typical applications include turbine blades, injection molds, and precision surgical instruments, whereas traditional machining methods are inadequate in accuracy and material integrity.

2) Laser beam machining: LBM is a high-precision, non-conventional machining process that uses a focused laser beam to remove material from the workpiece. This process is particularly effective for machining hard-to-cut materials such

as ceramics, superalloys, and composites, which are challenging to machine using traditional methods. LBM offers significant advantages in precision, flexibility, and the ability to machine complex shapes without direct contact with the workpiece.

With LBM, a high-energy laser beam is directed onto the surface of the workpiece. The intense energy of the laser beam heats, melts, and vaporizes the material in the target area. The material removal process is highly localized, minimizing the Heat-Affected Zone (HAZ) and reducing the risk of thermal deformation. Different lasers, such as CO2, Nd, and fiber, are used depending on the material and specific application requirements. Laser beam processing offers several notable advantages:

- Precision and accuracy: LBM can produce extremely fine features and intricate geometries with high precision and repeatability.
- Non-contact process: The lack of physical contact between the tool and the workpiece eliminates mechanical stresses and tool wear, enhancing the machining of brittle and hard materials.
- Minimal heat-affected zone: The localized nature of the laser beam minimizes thermal damage to surrounding material, maintaining the integrity and properties of the workpiece.
- Versatility: LBM can machine a wide range of materials, including metals, ceramics, polymers, and composites, making it suitable for various industrial applications.
- Despite its advantages, LBM also has some limitations:
- High initial investment: The cost of laser machining equipment can be significant, making it a considerable investment for manufacturers.
- Material-specific challenges: Different materials absorb laser energy differently, which can affect the efficiency and quality of the machining process. Optimizing parameters for each material is crucial.
- Thermal effects: Although minimal, there is still a risk of thermal effects such as micro-cracking or changes in material properties, particularly in sensitive materials.

Advances in laser technology have significantly improved the capabilities of LBM. High-power and ultra-short pulse lasers like femtosecond have improved material removal rates and machining precision. These lasers produce extremely short bursts of energy, minimizing heat diffusion and reducing the heat-affected zone. Beam delivery systems and control software advances have enabled more complex and precise machining operations. Due to its precision and versatility, LBM is widely used in various industries. Key applications include:

- Aerospace: Machining of turbine blades, intricate cooling channels, and other high-performance components made from superalloys and composites.
- Medical Devices: Manufacturing of precision surgical instruments, implants, and medical device components that require high accuracy and surface quality.
- Electronics: Fabrication of microelectronic components, Printed Circuit Boards (PCBs), and Micro-Electromechanical Systems (MEMS).
- Automotive: Production of precise components and tooling for high-performance engines and lightweight structures.

3) Ultrasonic machining: USM is an unconventional process that uses high-frequency ultrasonic vibrations to remove material from a workpiece. This technique is particularly effective for machining hard and brittle materials such as ceramics, glass, and advanced composites. USM offers significant advantages in precision, surface quality, and the ability to machine complicated shapes without substantial thermal or mechanical stress.

USM uses a tool vibrating at an ultrasonic frequency (typically 20 to 40 kHz) to strike abrasives onto the workpiece. The tool does not directly contact the workpiece but transfers energy through the abrasive particles suspended in the slurry. These particles act as cutting materials and erode the material from the workpiece through a combination of mechanical action and micro-chipping. The ultrasonic vibrations increase the effectiveness of the abrasive particles and lead to efficient material removal with minimal effort. USM offers several key advantages:

- Non-thermal process: USM does not generate significant heat during machining, eliminating thermal damage and preserving the workpiece's material properties.
- Minimal mechanical stress: The low amplitude vibrations used in USM induce minimal mechanical stress, reducing the risk of cracks and defects, especially in brittle materials.
- High precision and surface finish: USM can achieve high precision and excellent surface finishes, making it suitable for intricate and delicate components.
- Versatility: USM can machine various materials, including hard metals, ceramics, glass, and composites, offering broad applicability across different industries.

Despite its advantages, USM also faces some limitations:

- Slow material removal rate: Compared to other machining methods, USM has a relatively slow material removal rate, which can limit its application in high-volume production.
- Tool wear: The process's abrasive nature can lead to significant tool wear, necessitating frequent tool maintenance and replacement.

- Complex setup: The setup and maintenance of ultrasonic machining equipment require skilled operators and precise control systems.

Recent advances in ultrasonic machining have focused on improving process efficiency and expanding its applications. Innovations like multi-frequency ultrasound systems and hybrid machining processes have improved material removal rates and tool life. Hybrid processes that combine ultrasonic machining with other techniques, such as EDM or laser machining, leverage the strengths of both methods to achieve excellent results. Additionally, advancements in tool materials and abrasive slurries have improved performance and reduced wear. Due to its precision and ability to machine hard materials, ultrasonic machining is widely used in various industries. Key applications include:

- Aerospace: Machining of hard ceramic components, composite materials, and intricate features in high-performance parts.
- Medical devices: Manufacturing precision surgical instruments, dental tools, and medical implants that require high accuracy and smooth surfaces.
- Electronics: Fabrication of microelectronic components, optical devices, and advanced materials used in electronic packaging.
- Tool and die making: Production of complex molds, dies, and wear-resistant components from hard and brittle materials.

V. Conclusion

Precision machining of hard-to-cut materials remains a critical challenge and opportunity in modern manufacturing, driven by the need for high-performance components in the aerospace, automotive, biomedical, and electronics industries. This survey provided a comprehensive overview of the current state of precision machining techniques, focusing on conventional and non-conventional methods and highlighting advances in cutting tools, machining processes, and new technologies. Advances in cutting tool materials such as carbide, ceramic, PCD, and CBN have significantly improved the machinability of hard-to-cut materials by extending tool life, reducing wear, and enabling higher machining speeds. Innovative tool geometries and coatings also contribute to the efficiency and precision of the machining processes. Traditional machining methods, including turning and milling, have evolved with optimized cutting parameters, advanced cooling strategies, and real-time monitoring systems to meet the unique challenges of these materials. Meanwhile, unconventional machining techniques such as EDM, LBM, and USM offer alternative solutions for achieving high precision and complex geometries without causing thermal or mechanical damage.

Integrating digital technologies, including IoT, AI, and big data analytics, transforms precision machining and enables smarter, more efficient, and more adaptable manufacturing processes. These technologies improve process monitoring, predict tool wear, and optimize machining parameters in real time, improving productivity and quality. Future research and development should address the remaining challenges in precision machining, such as reducing tool wear, improving surface integrity, and developing more sustainable machining processes. Innovations in hybrid manufacturing processes and the continued development of digital and intelligent manufacturing technologies will play a critical role in overcoming these challenges and pushing the boundaries of what is possible when machining difficult-to-machine materials.

References

[1] J. Ahmad and Z. Zhou, "Mechanical properties of natural as well as synthetic fiber reinforced concrete: a review," Construction and Building Materials, vol. 333, p. 127353, 2022.

[2] S. B. Nagaraju, H. Priya, Y. G. T. Girijappa, and M. Puttegowda, "Lightweight and sustainable materials for aerospace applications," in Lightweight and Sustainable Composite Materials: Elsevier, 2023, pp. 157-178.

[3] H. T. Mohan, K. Jayanarayanan, and K. Mini, "A sustainable approach for the utilization of PPE biomedical waste in the construction sector," Engineering Science and Technology, an International Journal, vol. 32, p. 101060, 2022.

[4] M. Vishnupriyan, R. Annadurai, K. C. Onyelowe, and N. Ganasen, "Review on electronic waste used as construction materials-a scientometric analysis," Cogent Engineering, vol. 10, no. 2, p. 2283307, 2023.

[5] O. Adekomaya and T. Majozi, "Sustainable reclamation of synthetic materials as automotive parts replacement: effects of environmental response on natural fiber vulnerabilities," Environmental Science and Pollution Research, vol. 31, no. 12, pp. 18396-18411, 2024.

[6] R. Bag, A. Panda, A. K. Sahoo, and R. Kumar, "Sustainable high-speed hard machining of AISI 4340 steel under dry environment," Arabian Journal for Science and Engineering, vol. 48, no. 3, pp. 3073-3096, 2023.

[7] D. L. Soni and V. S. N. Neigapula, "Tool surface texturing in machining performance: state of art and recent developments," International Journal on Interactive Design and Manufacturing (IJIDeM), pp. 1-29, 2024.

[8] C. Zhang, D. Zou, M. Mazur, J. P. Mo, G. Li, and S. Ding, "The state of the art in machining additively manufactured titanium alloy Ti-6Al-4V," Materials, vol. 16, no. 7, p. 2583, 2023.

[9] E. García-Martínez, A. Molina-Yagüe, V. Miguel, and A. Martínez-Martínez, "Harmonic-based-on analysis to discriminate different mechanical actions involved in the machining of hard-to-cut materials," The International Journal of Advanced Manufacturing Technology, pp. 1-15, 2024.

[10] G. Gao, Y. Wang, Z. Fu, C. Zhao, D. Xiang, and B. Zhao, "Review of multi-dimensional ultrasonic vibration machining for aeronautical hard-to-cut materials," The International Journal of Advanced Manufacturing Technology, vol. 124, no. 3, pp. 681-707, 2023.

[11] I. Sharmin, M. A. Gafur, and N. R. Dhar, "Preparation and evaluation of a stable CNT-water based nano cutting fluid for machining hard-to-cut material," SN Applied Sciences, vol. 2, pp. 1-18, 2020.

[12] R. Kang, H. Ma, Z. Wang, Z. Dong, and Y. Bao, "Effect of tool wear on machining quality in milling Cf/SiC composites with PCD tool," Journal of Manufacturing Processes, vol. 105, pp. 370-385, 2023.

[13] R. Binali, H. Demirpolat, M. Kuntoğlu, and H. Sağlam, "Machinability investigations based on tool wear, surface roughness, cutting temperature, chip morphology and material removal rate during dry and MQL-assisted milling of Nimax mold steel," Lubricants, vol. 11, no. 3, p. 101, 2023.

[14] R. A. Laghari, M. Jamil, A. A. Laghari, A. M. Khan, S. S. Akhtar, and S. Mekid, "A critical review on tool wear mechanism and surface integrity aspects of SiCp/Al MMCs during turning: prospects and challenges," The

International Journal of Advanced Manufacturing Technology, vol. 126, no. 7, pp. 2825-2862, 2023.

[15] M. E. Korkmaz, M. Gupta, N. S. Ross, and V. Sivalingam, "Implementation of green cooling/lubrication strategies in metal cutting industries: A state of the art towards sustainable future and challenges," Sustainable Materials and Technologies, vol. 36, p. e00641, 2023.

[16] S. K. Lodhi, I. Hussain, and A. Y. Gill, "Artificial Intelligence: Pioneering the Future of Sustainable Cutting Tools in Smart Manufacturing," BIN: Bulletin of Informatics, vol. 2, no. 1, pp. 147-162, 2024.

[17] H. Attia et al., "Physics based models for characterization of machining performance–A critical review," CIRP Journal of Manufacturing Science and Technology, vol. 51, pp. 161-189, 2024.

[18] S. Seyedi, B. Pourghebleh, and N. Jafari Navimipour, "A new coplanar design of a 4 - bit ripple carry adder based on quantum - dot cellular automata technology," IET Circuits, Devices & Systems, vol. 16, no. 1, pp. 64-70, 2022.

[19] M. U. Farooq, M. A. Ali, S. Anwar, and H. A. Bhatti, "Process parameters optimization and performance analysis of micro-complex geometry machining on Ti6Al4V," International Journal on Interactive Design and Manufacturing (IJIDeM), pp. 1-21, 2024.

[20] M. Hourmand, A. A. Sarhan, M. Sayuti, and M. Hamdi, "A comprehensive review on machining of titanium alloys," Arabian Journal for Science and Engineering, vol. 46, pp. 7087-7123, 2021.

[21] D. Y. Pimenov et al., "A comprehensive review of machinability of difficult-to-machine alloys with advanced lubricating and cooling techniques," Tribology International, p. 109677, 2024.

[22] S. T. Haider, M. A. Shah, D.-G. Lee, and S. Hur, "A review of the recent applications of aluminum nitride-based piezoelectric devices," Ieee Access, vol. 11, pp. 58779-58795, 2023.

[23] N. Rathi, P. Kumar, and A. Gupta, "Non-conventional machining of nickel based superalloys: A review," Materials Today: Proceedings, 2023.

[24] Q. He et al., "Enhancing Tool Performance in High-Speed End Milling of Ti-6Al-4V Alloy: The Role of AlCrN PVD Coatings and Resistance to Chipping Wear," Journal of Manufacturing and Materials Processing, vol. 8, no. 2, p. 68, 2024.

[25] M. Sarıkaya et al., "A state-of-the-art review on tool wear and surface integrity characteristics in machining of superalloys," CIRP Journal of Manufacturing Science and Technology, vol. 35, pp. 624-658, 2021.

[26] H. Youssef, H. El-Hofy, A. Abdelaziz, and M. El-Hofy, "Accuracy and surface quality of abrasive waterjet machined CFRP composites," Journal of Composite Materials, vol. 55, no. 12, pp. 1693-1703, 2021.

[27] S. S. Akhtar, "A critical review on self-lubricating ceramic-composite cutting tools," Ceramics International, vol. 47, no. 15, pp. 20745-20767, 2021.

Analyzing VGG-19's Bias in Facial Beauty Prediction: Preference for Feminine Features

Nuno Fernandes[1], Sandra Soares[2], Joana Arantes[3]
School of Psychology, University of Minho, Braga, Portugal[1, 3]
Department of Education and Psychology, University of Aveiro, Aveiro, Portugal[2]

***Abstract*—From an evolutionary perspective, sexual dimorphism has been linked to perceived attractiveness, with masculine traits preferred in men and feminine traits in women. Moreover, symmetry is a strong predictor of facial attractiveness across both sexes. Recent advancements in the field of artificial intelligence have enabled algorithms to accurately predict facial attractiveness. This study aims to investigate whether these algorithms accurately replicate human judgments of attractiveness. We hypothesized that sexually dimorphic manipulations (masculinized men and feminized women) (H1), as well as symmetrized versions (H2), would elicit higher attractiveness ratings from a facial beauty prediction algorithm. Employing transfer learning, we trained six deep-learning models using four facial databases with attractiveness ratings (*n* = 6848). The top-performing model, VGG-19, demonstrated a high prediction correlation of .86 on the test set. Surprisingly, our findings revealed an interaction effect between sex and sexual dimorphism. Feminized versions of both men's and women's faces obtained higher attractiveness ratings than their masculinized counterparts. For symmetry, our results indicated that symmetrized faces were perceived as more attractive, albeit exclusively among women. These findings offer novel insights into the understanding of facial attractiveness from both algorithmic and human behavioral perspectives.**

***Keywords*—Deep learning; facial attractiveness; sexual dimorphism; symmetry; VGG-19**

I. Introduction

What makes a pretty face? The study of facial attractiveness is a multidisciplinary field that draws on knowledge from diverse disciplines, such as psychology, sociology, anthropology, and computer science [1-3]. Researchers in these disciplines employ diverse methods and approaches to explore the factors that contribute to facial attractiveness and its impact on social interactions and relationships.

In psychology, researchers investigate how people perceive and evaluate facial attractiveness, as well as the cognitive and neural mechanisms underlying this process [3]. They also explore the relationship between facial attractiveness and social cognition, such as the formation of impressions and romantic relationships [4-5]. Studies have shown that humans prefer associating with, dating, and mating with individuals considered facially attractive [6-8]. Additionally, attractive people tend to be perceived as more successful, enjoyable, and intelligent than unattractive people [4]. Specifically, research on human preferences for sexually dimorphic faces has garnered significant attention [9-10]. Sexual dimorphism in facial attractiveness refers to the physical differences in facial features that are considered more attractive in men and women. These differences in facial attractiveness are thought to be influenced by both evolutionary and cultural factors [see [11] for a comparison across five populations].

In sociology and anthropology, researchers explore how facial attractiveness and sexual dimorphism are related to cultural norms and values and how they shape social interactions regarding social status, power, and privilege [1]. Cross-cultural studies suggest a high consensus in facial attractiveness judgments across different populations [12-13]. However, research also indicates that cultural factors can shape perceptions of facial attractiveness, with different societies having distinct standards for what is considered attractive [14].

In computer science, computer vision researchers study how to create algorithms and models to predict facial attractiveness [2]. Explaining the functioning of deep learning models has recently gathered increasing attention from researchers and the wider public, including regulators and politicians [15-16]. While attractiveness prediction is of great use for the human-computer interaction field, it also raises significant moral and ethical questions. One of the primary concerns is whether beauty prediction algorithms reveal human-like psychophysical biases [17-18]. The study in [17] found that machine judgments of attractiveness displayed a preference for averaged faces and symmetrical faces, mirroring human judgments of attractiveness [19-20]. Additionally, the prediction of facial attractiveness holds clinical significance, particularly in the realm of plastic surgery [21-23]. To gain a comprehensive understanding of the complex issue of facial attractiveness, advancements in facial beauty prediction algorithms offer valuable insights. Furthermore, incorporating input from cognitive scientists in experimental design can enhance our understanding, especially in uncovering the decision-making processes behind algorithmic outputs.

A. Sexual Dimorphism and Attractiveness

From an evolutionary standpoint, males and females have evolved unique differences in their secondary sexual characteristics over time [24]. The development of more pronounced sexual dimorphic phenotypes – feminine traits for women and masculine traits for men – is considered attractive [9, 25], because it implies the inheritance of advantageous

genes favorable to offspring survival or reproductive success [26-27].

Evolutionary theories suggest that preferences for sexually dimorphic traits are associated with specific mating functions. For example, men may possess facial characteristics to indicate strength and dominance, such as an enlarged brow ridge, a thicker jawline, and a wider face [28]. Dominant men are considered more attractive by women [29-30], as they often achieve higher social status, thereby enhancing their capacity to provide essential resources for reproduction [31]. Additionally, masculinity is desirable as it signals higher testosterone levels [14, 32]. Since testosterone is immunosuppressive, compromising the body's ability to fight infections [33], only men with high genetic quality and robust immune systems can afford to invest in secondary sexual traits [34-35]. Therefore, men with more masculine traits are often perceived as healthier [36-37], and more attractive to potential mates [38-40].

Conversely, women may signal youth and fertility [41]. Feminine facial traits, such as a smaller jawline and a narrower face [26], indicate higher levels of estrogen, which correlate with superior reproductive qualities [42] and mating desirability [9, 25]. For instance, research has shown that both men and women rated photographs of women's faces captured during the fertile window of the menstrual cycle (late follicular) as more appealing compared to photographs taken during the non-fertile (luteal) phase [43]. Similarly, [44] explored the impact of facial shape transformations towards late follicular and luteal prototypes on male perceptions of attractiveness. Their findings revealed a distinct preference for faces resembling the late follicular phase, suggesting that subtle shape differences can sway men's preferences depending on a woman's menstrual cycle phase. During the late follicular period, a peak of estrogen and luteinizing hormone (LH) leads to ovulation and is followed by a rise in progesterone in the luteal phase [45-47]. This increased attractiveness may serve as an adaptive mechanism to enhance a female's perceived value in the mating pool during the phase of the cycle when the likelihood of conception is at its peak [43]. Ultimately, the preference for women's feminine facial features remains robust across the literature [10, 48-52].

B. Algorithms and Attractiveness

Early approaches to quantifying facial attractiveness were grounded in the notion that specific ratios and proportions of facial features are more appealing [17, 53-54] For example, the "golden ratio" or "phi" (1.618) has been used to measure facial proportions and attractiveness [56]. The idea is that if specific facial ratios, such as the distance between the eyes and the distance between the mouth and eyes, conform to phi, the face is more attractive. However, contemporary computer-based methods, including machine learning and deep learning algorithms, have shown more promising outcomes.

Machine learning offers a method that can be used to predict facial attractiveness by analyzing patterns in predefined facial features known to be associated with attractiveness [17, 53-54]. These models assess multiple hand-crafted features, such as symmetry, geometric ratios, and distances between landmarks. Another approach uses deep learning algorithms, such as Convolutional Neural Networks (CNNs), which also predict facial attractiveness. One advantage of CNNs is that they do not require pre-defined facial attributes, as they can extract useful features directly from raw images. Consequently, these models have achieved a relatively high correlation with the actual judgments of attractiveness [56-58].

Furthermore, leveraging transfer learning eliminates the need to develop an algorithm from the ground up, as it allows for the use of pre-existing models trained on large datasets, such as architectures from the Visual Geometry Group Network (VGG; [59]) or residual networks (ResNet) ([60]). Although these models were initially designed for image classification tasks, they can be fine-tuned for facial beauty prediction [61-63]. The fine-tuning process involves adjusting the model's parameters to better fit the new task, typically by training the final layers of the model on the new dataset. Once the model is fine-tuned, it can predict the attractiveness of new facial images. This approach has been widely used in attractiveness prediction tasks as it saves computational time and resources while improving performance [61-63].

The benchmark for training and evaluating facial beauty prediction algorithms is the SCUT-FBP5500 [64], with cutting-edge models achieving a Pearson correlation of .93 with human judgments of attractiveness [57]. However, despite the large number of images (n = 5500), attractiveness ratings were performed only by Asian raters. Therefore, this could be a potential limitation for generalizing the beauty prediction algorithms. Additionally, the high accuracy observed could be attributed to greater internal consistency in attractiveness judgments within the Asian population [10].

C. The Present Study

Despite significant advancements in facial beauty prediction algorithms, a gap remains in the literature regarding how these algorithms make their decisions. The main objective of this study is to investigate whether beauty prediction algorithms replicate human biases toward sexually dimorphic facial traits. We hypothesize that sexually dimorphic versions of male (masculinized) and female (feminized) facial images will be perceived as more attractive (H1). Moreover, we will manipulate symmetry, which is a well-established predictor of facial attractiveness [19,65-67]. We expect that symmetrized facial images of both men and women will be rated as more attractive (H2). Our secondary aim is to provide a benchmark for facial attractiveness prediction using multiple databases. This novel approach aims to enhance our understanding of how facial beauty prediction algorithms assess attractiveness and whether they support the sexual dimorphism and symmetry hypotheses.

II. METHOD

A. Materials

1) Databases: For the transfer learning phase, four databases were used: SCUT-FBP5500 [64]; Chicago Face Database (CFD) [68]; Karolinska Directed Emotional Faces (KDEF) [69]; and FACES [70], resulting in a total of 6,848 images. Attractiveness ratings across all datasets were

normalized to a scale from 0 to 1. An additional database, the Face Research Lab London (FRL-London; [71]), was used to create sexually dimorphic and symmetrized versions of faces, which were then evaluated for facial attractiveness by the trained model. Following common preprocessing practices for image data in deep learning tasks [59-60] pixel values were rescaled to the range of (0, 1) by dividing each pixel intensity value by 255. This normalization step ensures that the input features are appropriately scaled and facilitates efficient training of machine learning models, particularly neural networks, by mitigating issues such as vanishing or exploding gradients [72-73].

The SCUT-FBP5500 [64] dataset consists of 5500 frontal face images, including 2000 Asian females, 2000 Asian males, 750 Caucasian females, and 750 Caucasian males. Most of the faces have neutral expressions and simple backgrounds. The faces in the dataset were labeled with attractiveness ratings from 1 to 5 by 60 Asian raters aged 15 to 60.

The Chicago Face Database (CFD) [68] consists of 970 facial images displaying neutral, happy, threatening, and fearful expressions. The 158 volunteers who photographed included 37 Black males, 48 Black females, 36 White males, and 37 White females, ranging in age from 18 to 40 years. A sample of 1,087 raters made subjective ratings of the image's attractiveness on a 1–7 Likert scale, with each subject rating only 15 faces. Participants included 552 females, 308 males, and 227 who preferred not to disclose their gender, with an average age of 26.75 (SD_{age} = 10.54). They came from diverse racial backgrounds: 516 White, 117 Asian, 74 Black, 72 biracial or multiracial, 57 Latino, 18 other, and 233 who did not report their race.

The Karolinska Directed Emotional Faces (KDEF) [69] is one of the most widely used databases of human facial expressions. [74] presented subjective attractiveness ratings for a subset of 210 pictures from 70 Caucasian amateur actors (35 women and 35 men, aged between 20 and 30 years) with different facial expressions: angry, happy, and neutral. The sample of raters included 155 students from Portuguese universities (83.20% female; M_{age} = 23.73; SD_{age} = 7.24). Each participant rated the facial attractiveness of 36 pictures on a 1-7 Likert scale.

The FACES [70] is a set of facial images from 171 women and men categorized into three age groups: young (n = 58; age range: 19–31; M_{age} = 24.3; SD_{age} = 3.5), middle-aged (n = 56; age range: 39-55; M_{age} = 49.0; SD_{age} = 3.9), and older (n = 57; age range: 69-80; M_{age} = 73.2; SD_{age} = 2.8). Each individual displays six facial expressions: neutrality, sadness, disgust, fear, anger, and happiness, resulting in a total of 1026 pictures. Facial attractiveness evaluations were performed by 154 raters, all of whom were Caucasian and German. Each participant was randomly assigned to either set A or set B, which consisted of identical pictures of the same subjects and facial expressions, with only minor differences (e.g., head inclination angle). For the present work, only set A was considered. Each image was rated by a minimum of 8 and a maximum of 14 raters per age group by gender. Therefore, six attractiveness ratings were provided in the metadata of the original work. In our case, we used the average of these ratings on a scale from 1 to 100.

The Face Research Lab London (FRL-London) was chosen for the sexual dimorphic transformations [71], which consists of 102 neutral front faces (male = 52; female = 50; M_{age} = 26.9; SD_{age} = 7.07) from different ethnicities (69 White, 13 Black, and 20 Asian). We used only 100 photos from this database due to problems manipulating two male images (codes 005_03 Asian and 114_03 Black), resulting in an equal number of male and female facial images. This dataset included pre-delineated face shape templates of 189 coordinates for each image. Attractiveness ratings (on a 1-7 scale) were made by 2,513 people (age range: 17-90).

2) Algorithms: A set of six facial recognition models was used to determine which of these models would be the most adequate for facial attractiveness prediction. The selected models were: InceptionResNetV2 [75], MobileNetV2 [76], EfficientNetV2B0 [77], ResNet50 [60], Xception [78], and VGG-19 [59].

B. Procedure

The procedure was divided into three phases. Initially, we employed a transfer learning methodology across a range of models, followed by fine-tuning the one that yielded the most favorable results. In the second phase, we applied sexual dimorphic and symmetrical transformations of the images from the FRL-London dataset for further evaluation. Finally, we utilized the fine-tuned model to predict the facial attractiveness of the original, sexually dimorphic, and symmetrized versions of both female and male facial photos.

1) Transfer learning: We selected a set of six commonly used architectures with pre-trained weights for image classification problems from the Keras Applications Module: tf.keras.applications [79]. Then, we applied transfer learning to each model to predict facial attractiveness, similar to previous studies [61-63]). This involved freezing all layers except the last one, and transforming the initial classification problem into a regression problem by incorporating a final dense layer.

The models were trained for 100 epochs with patience set to 30 epochs, using the Adam optimizer with a learning rate of 0.001 and weight decay of 0.004. Adam is an "algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments" [80]. A train/validation/test split of 60/20/20 size was used, resulting in 4110/1369/1369 images for the corresponding set. The model with the best results was the VGG-19 with a Pearson's correlation of .76 (*RMSE* = 0.094; *MAE* = 0.07).

The VGG-19 belongs to a series of deep neural networks from the Visual Geometry Group Network (VGG), also including VGG-11, VGG-13, and VGG-16 ([59]). These networks share their structure by having several blocks of convolutional layers connected to a final block consisted of three fully connected layers. The VGGNet has been trained on over one million images in 1000 classes [59]. Specifically,

VGG-19 has five blocks of convolution layers and one last block of three fully connected layers. A 3x3 Max Pooling layer connects all blocks. The first (size 64x64x64) and second (size 128x128x128) blocks have two convolutional layers, while the third (size 256x256x256), fourth (size 512x512x512) and fifth (size 512x512x512) blocks have four convolutional layers. Since VGG-19 showed the best performance for predicting facial attractiveness compared to the other architectures, we fine-tuned this model.

Therefore, we unfroze the 3-fully connected layers, whereas in the previous phase only the last regression layer was trainable. We used the same hyperparameters and train/validation/test sets. The model's performance on the test set increased from a Pearson's correlation of .76 to .84 (*RMSE* = 0.12; *MAE* = 0.10). Finally, we fine-tuned the model with the best parameters by unfreezing all the model's layers. We set the training for 10 epochs (patience of 5), using Adam optimizer with a learning rate of 0.00001 and weight decay of 0.004, to prevent overfitting [see [81] for a revision on hyperparameter optimization for fine-tuning]. A final correlation of .86 ($p < .001$) was achieved between the model's predictions and the test set (*RMSE* = 0.10; *MAE* = 0.08).

2) Image preprocessing: To achieve sexual dimorphic and symmetric transformations, we used the FRL-London [71] database. All image manipulations were performed using computer vision techniques provided by the Psychomorph software [82] and its web-based version, WebMorph [83]. Following previous research [52, 84-85], sexually dimorphic versions of each image were created using a 50% spatial linear transformation towards either a female avatar (feminized version) or a male avatar (masculinized version). The avatars (average faces) were generated by averaging 30 male faces (male avatar) and 30 female faces (female avatar). Additional information regarding theses methods can be found in [86]. Symmetrized versions of both the original and manipulated photos (feminized and masculinized faces) were created by mirror reversing the image and then combining one side of the original non-symmetrized photo with the opposite side of the reversed photo [67, 87].

3) Data analysis: The algorithm's prediction of attractiveness was used as the dependent variable in a linear mixed-effects model (LME) with three fixed main effects: sexual dimorphism manipulation (masculinized/feminized), symmetry (original/symmetric), and sex (male/female), along with the three-way interaction term. The photo's code was included as a random factor with varying intercepts to account for individual differences in attractiveness, as the model predicted attractiveness for multiple versions (sexually dimorphic and symmetric transformations) of the same individual [see [88] for a detailed explanation on LME models]. A post-hoc analysis using Bonferroni adjustment was performed for the observed significant effects. All statistical analyses were conducted using R Statistical Software (v4.3.1; [89]) using lme4 (v.1.1.34; [90]) and ggplot2 (v.3.4.4.9000; [91]) for plotting. The data and reproducible code are publicly available at [92].

III. RESULTS

The LME (estimated using REML) was composed of both main effects (sexual dimorphism manipulation, symmetry, and sex) and its three-way interaction term (Fig. 1). The model included the image's code as a random factor. The model's total explanatory power was substantial (conditional $R^2 = .93$), and the part related to the fixed effects alone (marginal R^2) was equal to .33. The model showed a statistically significant main effect of sexual dimorphism, $F(2, 490) = 64.02$, $p < .001$, symmetry, $F(2, 490) = 32.07$, $p < .001$, and sex, $F(1,98) = 49.74$, $p < .001$. For the interaction terms, the two-way interaction of sexual dimorphism by symmetry manipulation was shown to be statistically significant, $F(2,490) = 5.40$, $p = .005$, while the two-way interaction of sexual dimorphism and sex, $F(2,490) = 2.73$, $p = .066$, or the triple interaction of sexual dimorphism*symmetry*sex, $F(2,490) = 0.91$, $p = .40$, showed no significant results.

Post-hoc analysis using Bonferroni correction showed that feminized faces were rated as more attractive for both male and female images (H1). In the non-symmetric condition, feminized females were judged as more attractive than masculinized females, $t(490) = 4.41$, $p < .001$, 95% CI [0.01, 0.02], $d = 0.88$, as well as in the symmetric condition, $t(490) = 5.44$, $p < .001$, 95% CI [0.01 0.02], $d = 1.09$. Also, feminized males in the non-symmetric condition were rated as more attractive than masculinized ones, $t(490) = 5.47$, $p < .001$, 95% CI [0.01, 0.02], $d = 1.09$, as well as in the symmetric condition, $t(490) = 6.34$, $p < .001$, 95% CI [0.01 0.03], $d = 1.27$. Additionally, feminized manipulations were perceived as more attractive than the original faces for all conditions, except for females in the non-symmetrized version, $t(490) = 0.77$, $p = 1$, 95% CI [-0.01, 0.01], $d = 0.15$. Similarly, no differences were found between the masculinized versions and the original faces, except for that condition in which masculinized female photos were judged as less attractive than the original ones in the non-symmetrized versions, $t(490) = -3.64$, $p = .007$, 95% CI[-0.02, -0.01], $d = -0.73$.

For the symmetry main effect that was observed (H2), post-hocs revelated that symmetrized versions were only more attractive for feminized female faces, $t(490) = 4.43$, $p < .001$, 95% CI [-0.02, -0.01], $d = 0.89$, and masculinized female faces $t(490) = -3.41$, $p = .02$, 95% CI [-0.02, -0.01], $d = 0.58$, in comparison with the non-symmetrized ones. No differences were observed for males or between original photos.

Finally, supporting the sex main effect, in the non-symmetrized condition, feminized females were perceived as more attractive than feminized males, $t(118) = 6.12$, $p < .001$, 95% CI [0.03, -0.08], $d = 3.71$, with the same occurring for masculinized versions, $t(118) = 6.47$, $p < .001$, 95% CI [0.03, 0.08], $d = 3.92$, and original faces, $t(118) = 7.28$, $p < .001$, 95% CI [0.04, 0.09], $d = 4.42$. In the symmetrized condition, the same pattern was found, with feminized female photos being rated as more attractive than feminized male photos, $t(118) = 6.61$, $p < .001$, 95% CI [0.03, 0.08], $d = 4.01$, as well as masculinized versions, $t(118) = 6.90$, $p < .001$, 95% CI

[0.03, -0.09], d = 4.19, and original ones, $t(118) = 6.98$, $p < .001$, 95% CI [0.03, 0.09], d = 4.23.

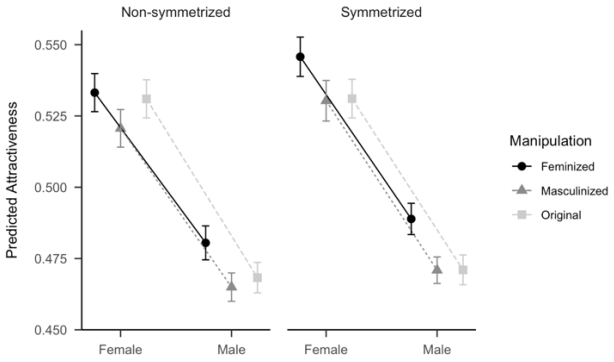


Fig. 1. The attractiveness ratings of VGG-19.

IV. DISCUSSION

The main aim of the present study was to investigate whether beauty prediction algorithms replicate human judgments concerning facial sexually dimorphic traits. We hypothesized that sexually dimorphic versions of male (masculinized) and female (feminized) facial pictures would be rated as more attractive (H1). Also, we expected that symmetrized versions of both male and female facial pictures would be perceived as more attractive (H2). Additionally, we aimed to provide a benchmark for facial attractiveness prediction using multiple databases.

The fine-tuned model was used to predict the attractiveness of symmetrized and non-symmetrized versions of masculinized, feminized, and original frontal facial images with neutral expressions from the FRL-London database. Surprisingly, our results showed that feminized manipulations were rated as more attractive compared to the masculinized versions for both male and female faces. Thus, we partially confirmed our first hypothesis (H1), as we expected a sexual dimorphism main effect, with feminized versions being preferred for females, but not for males. Moreover, despite the observed main effect of symmetry, post-hoc analyses suggested that symmetry increased facial attractiveness specifically for feminized and masculinized versions of female faces. Thus, the data also partially supported our second hypothesis (H2). Additionally, across all conditions, females were consistently rated as more attractive than males.

We achieved a Pearson correlation of .86 with the fine-tuned VGG-19 on the test set across four databases. Although slightly lower than state-of-the-art results (e.g., [57] with r = .93), our approach addresses a key limitation by reducing bias from relying solely on the SCUT-FBP5500 dataset. This offers a valuable benchmark for facial beauty prediction, encouraging future research to explore and improve on these findings.

A. Theoretical Implications

Firstly, our work contributes to understanding how facial beauty prediction algorithms make decisions. The consistent overall preference for feminine facial traits suggests that beauty prediction algorithms can implicitly learn human attractiveness preferences, like [17] findings. More recently, [18], using the VGG architecture, showed that DNNs use putative ratios (e.g., golden ratio) as essential attributes for predicting facial attractiveness, akin to human judgments. This convergence between algorithmic and human judgments underscores the complex interplay between facial features, attractiveness perception, and evolutionary pressures. Further investigation into the mechanisms underlying both facial attractiveness prediction and sexual dimorphism is therefore warranted.

1) Sexual dimorphism: Regarding our first hypothesis (H1), and despite our initial assumptions, feminized versions of both male and female faces were rated as more attractive. The preference for femininity in women is well-supported in the existing body of literature [10, 48, 50, 93-94]. However, there are mixed findings concerning the effect of sexual dimorphism on male attractiveness. Some research reports a stronger preference for more masculine male faces [95-100], while other studies suggest a preference for more feminine faces [10,37,97,101-104], or report no differences between masculinized and feminized versions of male faces [50,105-106].

A possible explanation for these diverse findings may be associated with the level of attractiveness from both the rater and the photo. For instance, [107] found that when judging attractive faces, both male and female participants favored masculine male faces over feminine ones. Conversely, in the less attractive condition, they preferred feminine male faces to masculine ones. Furthermore, [108] found that women's preferences for femininity in men's faces decreased after viewing images of highly attractive men. Since the predicted attractiveness in our results may be considered average, this could explain the lack of preference for masculinity. Moreover, there is an increased preference for masculinity among women who perceive themselves as attractive [102]. Furthermore, women perceived as having relatively low facial attractiveness by others, showed a preference for more "feminine" male faces when selecting partners for long-term relationships compared to short-term relationships [85].

Furthermore, enhancing masculine facial features may increase perceptions of dominance and negative attributes, such as coldness or dishonesty [10]. These aspects are considered significant by women within the context of relationships and paternal investment [10]. Subsequent research has unveiled that women prefer less masculine male faces in long-term relationships under specific conditions of environmental harshness, such as resource scarcity or elevated stress levels [39,109]. These findings align with the notion that men with more favorable genes (higher testosterone levels) invest less in the relationship, as they can afford to choose their mate [109].

However, it is important to note that this aspect remains ambiguous within the existing body of literature. In a recent study, women exhibited a preference for masculinized faces over feminized ones [100]. Notably, this preference for masculine male faces was more pronounced when assessing potential co-parenting partners compared to short-term

relationships [100]. Supporting these findings, [110] demonstrated that facial attractiveness, instead of facial masculinity influenced perceptions of paternal involvement.

Finally, hormonal changes occurring during the menstrual cycle may alter women's preference for sexually dimorphic cues. In women, the preference for masculinity is higher during the late follicular phase of the menstrual cycle (when fertility is high and progesterone level low) [111]. Similarly, [112] found that women tend to prefer masculinity in male faces around ovulation, while they prefer feminized male faces during the rest of the cycle. This likely indicates a balanced trade-off between the attraction to males who appear to ensure reproductive success and males who are perceived as capable and nurturing fathers.

2) Symmetry: Concerning our second hypothesis (H2), our results showed that the facial beauty prediction algorithm found symmetrized faces more attractive, but only for sexually dimorphic versions of female faces. Similar to our findings, machine learning models trained to predict facial attractiveness based on geometric facial features have been found to implicitly learn human psychophysical biases towards symmetric faces [53,113]. However, it is important to note that the specific characteristics that make symmetric faces more attractive, independent of their symmetry, are not definitively defined.

Since our results showed a preference for symmetric sex-dimorphic traits exclusively in female faces, insights from prior studies [97,114] suggesting an intercorrelation between symmetry and sexual dimorphism may shed light on why symmetrized male faces were not found more attractive than non-symmetrized versions. In a cross-cultural study, perceived sexual dimorphism demonstrated a correlation with symmetry measurements [97]. Additionally, research found that women's preferences for symmetry were positively correlated with preferences for masculinity in male faces and that men's preferences for symmetry were positively correlated with preferences for femininity in female faces [97]. Furthermore, [115] found that facial symmetry was associated with masculine facial cues in males, such as increased lower face length and cheekbone prominence [32,116]. Symmetry could potentially enhance the perception of masculinity or be interpreted by the algorithm as a cue of masculinity in male faces, which might explain why the effect was observed only for female faces.

Despite the critical role of symmetry in the perception of attractiveness, evidence suggests that sexual dimorphic cues may be more influential than symmetry in attractiveness evaluations, which supports our findings. For instance, [117] found that participants focused more on sexual dimorphic cues than on symmetry when selecting a partner for a romantic relationship. Furthermore, [118] altered symmetry within a face while keeping the mean size of facial features constant, revealing that symmetric faces were perceived as less attractive. Additionally, [119] found that symmetry enhances facial attractiveness by increasing perceived normality, suggesting that attractiveness perception may involve additional inferences about a face. Likewise, using a random forest machine learning algorithm to predict facial attractiveness, [20] found that shape averageness, dimorphism, and skin texture symmetry emerged as valuable features capable of yielding relatively precise predictions. However, it was observed that shape symmetry did not contribute significantly to the predictive accuracy.

3) Sex: Our findings indicate that the facial beauty prediction algorithm rated women as more attractive than men. This observation aligns with previous research that utilized neural networks with facial landmarks as input data [120], as well as studies based on human assessments of attractiveness [121-122]. Considering the sex effect observed, it could be argued that the algorithm's general preference for feminization may stem from an inclination towards feminine traits, regardless of the sex of the face being evaluated. However, several studies have highlighted the impressive capability of convolutional neural networks in sex recognition tasks, e.g., [123-124].

B. Limitations / Further Studies

There are some limitations mainly associated with the databases that warrant consideration. Novel research has revealed that beauty prediction algorithms trained using SCUT-FBP5500 images exhibit poor prediction accuracy when assessing faces from other databases [125]. There may be a cultural effect in our study, as the largest database consisted solely of Asian raters, potentially influencing preferences for masculinity in male faces. Research suggests that a preference for more masculine men is more pronounced in countries with higher economic and national health indices [103]. To prevent the model from overfitting to the characteristics of the SCUT-FBP5500 dataset, we included additional datasets with standardized frontal facial images.

Furthermore, there is an interrelationship between attractiveness and emotional expression [74,126,128]. Human raters consistently perceive attractive faces as more appealing [75,126], while positive emotions are more readily recognized in attractive faces [126-127]. While emotion recognition algorithms have advanced [128], they haven't yet integrated attractiveness ratings, and beauty prediction models typically overlook emotionality. Future research should explore whether algorithms replicate human biases regarding attractiveness and emotions, and examine their interaction with sexual dimorphism.

Additionally, integrating multiple databases with varying rating scales for facial beauty prediction can pose challenges, as noted by [125]. [129] addresses the challenge of standardizing Likert scales, noting that the mean may shift depending on the scale's point range. To address this issue, [129] suggests a rescaling system based on transitional probabilities. However, given that the scales from the selected datasets range from 1-5 to 1-100, we opted for a straightforward normalization to the 0-1 range. Nonetheless, using ordinal attractiveness judgments as scalar intervals presents a rescaling challenge. The interval properties of Likert scales are inherent to the data, influenced by observers' mean responses rather than the labels themselves [130]. Thus,

we advocate for a unified rating system for attractiveness datasets to enhance predictive models.

In this study, VGG-19 was trained using attractiveness ratings from both male and female evaluators who assessed photos of individuals from both the same and opposite sexes. From an evolutionary perspective, the sexual dimorphism hypothesis posits that individuals seek high-quality partners for reproduction. Therefore, the same-sex attractiveness judgments in these datasets may represent a potential limitation. However, as previously discussed, [107] demonstrated congruent attractiveness preferences among men and women, indicating a preference for masculinity in attractive male faces, whereas feminine traits were preferred in less attractive male faces. Additionally, earlier research, such as that by [131], did not find differences between same-sex and opposite-sex attractiveness preferences. Supporting the preference for sexual dimorphism, [81] reported that male participants rated masculinized versions of male faces as more attractive than feminine faces. Moreover, the self-rated sex-typicality appears to moderate this effect. For instance, individuals who perceive themselves as possessing more exaggerated sex-typical traits tend to exhibit stronger preferences for exaggerated sex-typical shape cues in faces of the same sex [132]. Further studies may explore whether same-sex or opposite-sex ratings could explain the observable effects. Therefore, researchers should consider incorporating this factor when developing future databases of facial attractiveness.

Furthermore, sexual orientation may influence preferences for sex-dimorphic face shapes. Research indicates that homosexual men tend to have a stronger preference for masculinity in male faces compared to other demographic groups [133]. However, there is evidence suggesting that sexual position moderates facial masculinity preferences [134-135]. Individuals' self-labeling as "top" or "bottom", as well as their preferences for their partner's sexual position, are indicative of their preferences for masculinity. Generally, individuals seeking a "top" as a romantic partner prefer more masculine faces than those seeking "bottoms" [134-135]. Additionally, homosexual women exhibited stronger preferences for masculinity in female faces compared to heterosexual women [133]. Thus, sexual orientation may be a crucial factor to consider in subsequent models. To account for this, we suggest that authors incorporate sexual orientation when developing databases of facial attractiveness and make this information available to other researchers.

V. Conclusion

The results of the present study align with previous findings in the psychological literature, which indicate a preference for femininity in both sexes. Moreover, we supported prior findings about a preference for symmetric faces. However, our results reveal an interplay between sexual dimorphism and symmetry, as the preference for symmetric faces was observed only in dimorphic faces, not in the original versions. Furthermore, our results provide a benchmark for developing beauty prediction algorithms using multiple databases. Future research could build on these findings by investigating additional cues of facial attractiveness. Overall, our findings suggest that beauty prediction algorithms can implicitly learn and replicate human biases regarding facial attractiveness.

Acknowledgment

This study was conducted at the Psychology Research Centre (CIPsi), School of Psychology, University of Minho, supported by the Foundation for Science and Technology (FCT) through the Portuguese State Budget (Ref.: UIDB/PSI/01662/2020). This work was also supported by the Portuguese Foundation for Science and Technology (FCT) through the individual research grant 2021.05008.BD.

References

[1] Frevert, T. K., & Walker, L. S. (2014). Physical attractiveness and social status. *Sociology Compass, 8*(3), 313–323.

[2] 2d, J. N., & Abdulazeez, A. M. (2021). Facial beauty prediction and analysis based on deep convolutional neural network: A Review. *Journal of Soft Computing and Data Mining, 2*(1), 1–12.

[3] Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences, 3*(12), 452-460.

[4] Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*(3), 285–290.

[5] Talamas, S. N., Mavor, K. I., & Perrett, D. I. (2016). Blinded by beauty: Attractiveness bias and accurate perceptions of academic performance. *PLoS ONE 11*(2): e0148284.

[6] Buss, D. M. (1994). *The evolution of desire: Strategies of human mating*. Basic Books.

[7] Rhodes, G., Simmons, L. W., & Peters, M. (2005). Attractiveness and sexual behavior: Does attractiveness enhance mating success?. *Evolution and Human Behavior, 26*(2), 186-201.

[8] Riggio R., & Woll S. (1984). The role of non-verbal and physical attractiveness in the selection of dating partners. *Journal of Social and Personal Relationships, 1*(3), 347–357.

[9] Barber, N. (1995). The evolutionary psychology of physical attractiveness: Sexual selection and human morphology. *Ethology and Sociobiology, 16*(5), 395–424.

[10] Perrett, D., Lee, K., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., … & Akamatsu, S. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature, 394*, 884–887.

[11] Fiala, V., Třebický, V., Pazhoohi, F., Leongómez, J. D., Tureček, P., Saribay, S., … & Kleisner, K. (2021). Facial attractiveness and preference of sexual dimorphism: A comparison across five populations. *Evolutionary Human Sciences, 3*, E38.

[12] Cunningham, M. R., Roberts, A. R., Barbee, A. P., Druen, P. B., & Wu, C.-H. (1995). "Their ideas of beauty are, on the whole, the same as ours": Consistency and variability in the cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology, 68*(2), 261–279.

[13] Jones, D., & Hill, K. (1993). Criteria of facial attractiveness in five populations. *Human Nature, 4*, 271-296.

[14] Penton-Voak, I. S., Jacobson, A., & Trivers, R. (2004). Populational differences in attractiveness judgements of male and female faces: Comparing British and Jamaican samples. *Evolution and Human Behavior, 25*(6), 355-370.

[15] Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: an analytical review. *Data Mining and Knowledge Discovery*, *11*(5), e1424.

[16] Samek, W., & Müller, K. R. (2019). Towards Explainable Artificial Intelligence. In: W. Samek, G. Montavon, A. Vedaldi, L. Hansen, K. R. Müller (Eds.) *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*(pp. 5-22). Springer.

[17] Kagian, A., Dror, G., Leyvand, T., Meilijson, I., Cohen-Or, D., & Ruppin, E. (2008). A machine learning predictor of facial attractiveness revealing human-like psychophysical biases. *Vision Research, 48*(2), 235-243.

[18] Tong, S., Liang, X., Kumada, T., & Iwaki, S. (2021). Putative ratios of facial attractiveness in a deep neural network. *Vision Research, 178*, 86-99.

[19] Grammer, K., & Thornhill, R. (1994). Human (Homo sapiens) facial attractiveness and sexual selection: The role of symmetry and averageness. *Journal of Comparative Psychology*, *108*(3), 233–242.

[20] Jones, A. L., & Jaeger, B. (2019). Biological bases of beauty revisited: The effect of symmetry, averageness, and sexual dimorphism on female facial attractiveness. *Symmetry, 11*(2), 279.

[21] Kanevsky, J., Corban, J., Gaster, R., Kanevsky, A., Lin, S., & Gilardino, M. (2016). Big data and machine learning in plastic surgery: A new frontier in surgical innovation. *Plastic Reconstructive Surgery, 137*(5), 890e-897e.

[22] Štěpánek, L., Měšťák, J., & Kasal, P. (2019). *Machine-learning at the service of plastic surgery: a case study evaluating facial attractiveness and emotions using R language*. 2019 Federated Conference on Computer Science and Information Systems (FedCSIS), Leipzig, Germany, 2019, pp. 107-112,

[23] Štěpánek, L., Kasal, P., & Měšťák, J. (2019). *Evaluation of Facial Attractiveness after Undergoing Rhinoplasty Using Tree-based and Regression Methods*. 2019 E-Health and Bioengineering Conference (EHB), Iasi, Romania, pp. 1-4.

[24] Darwin, C. (1871). The Descent of Man, and Selection in Relation to Sex. John Murray.

[25] Symons, D. (1995). Beauty is in the adaptations of the beholder: The evolutionary psychology of human female sexual attractiveness. In P. R. Abramson & S. D. Pinkerton (Eds.), Sexual nature, sexual culture (pp. 80–119). The University of Chicago Press.

[26] Little A. C., Jones B. C., & DeBruine L. M. (2011). Facial attractiveness: evolutionary based research. *Philosophical transactions of the Royal Society of London*, 366(1571), 1638-59.

[27] Puts, D. A. (2010). Beauty and the beast: Mechanisms of sexual selection in humans. *Evolution and Human Behavior*, *31*(3), 157–175.

[28] Scott, I. M. L., Clark, A. P., Boothroyd, L. G., & Penton-Voak, I. S. (2013). Do men's faces really signal heritable immunocompetence? *Behavioral Ecology 24*(3), 579–589.

[29] Neave, N., Laing, S., Fink, B., & Manning, J. T. (2003). Second to Fourth Digit Ratio, Testosterone and Perceived Male Dominance. *Proceedings: Biological Sciences, 270*(1529), 2167–2172.

[30] Townsend, J. M., & Roberts, L. W. (1993). Gender differences in mate preference among law students: Divergence and convergence of criteria. *The Journal of Psychology: Interdisciplinary and Applied, 127*(5), 507–528.

[31] Geary, D. C., Vigil, J., & Byrd-Craven, J. (2004). Evolution of human mate choice. *The Journal of Sex Research, 41*(1), 27-42.

[32] Pound. N., Penton-Voak, I. S., & Surridge, A. K. (2009). Testosterone responses to competition in men are related to facial masculinity. *Proceedings of the Royal Society of London, B,* 276(1654), 153-9.

[33] Foo, Y. Z, Nakagawa, S., Rhodes, G., & Simmons, L. W. (2017). The effects of sex hormones on immune function: a meta-analysis. *Biological Reviews, 92*(1), 551-571.

[34] Folstad, I., & Karter, A. J. (1992). Parasites, Bright Males, and the Immunocompetence Handicap. *The American Naturalist, 139*(3), 603–622.

[35] Hamilton, W. D., & Zuk, M. (1982). Heritable true fitness and bright birds: A role for parasites? *Science, 218*(4570), 384–387.

[36] Dixson, B. J., & Brooks, R. C. (2013). The role of facial hair in women's perceptions of men's attractiveness, health, masculinity and parenting abilities. *Evolution and Human Behavior, 34*(3), 236-241.

[37] Rhodes G., Chan J., Zebrowitz, L. A., & Simmons, L. W. (2003). Does sexual dimorphism in human faces signal health? *Proceedings of the Royal Society of London, Biological Sciences, 270*(suppl_1), S93–S95.

[38] Little, A. C., Jones, B. C., & Burriss, R. P. (2007). Preferences for masculinity in male bodies change across the menstrual cycle. *Hormones and Behavior, 51*(5), 633-639.

[39] Little, A. C., Cohen, D. L., Jones, B. C., & Belsky, J. (2007). Human preferences for facial masculinity change with relationship type and environmental harshness. *Behavioral Ecology and Sociobiology, 61*, 967–973 (2007).

[40] Rennels, J. L., Bronstad, P. M., & Langlois, J. H. (2008). Are attractive men's faces masculine or feminine? The importance of type of facial stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 34*(4), 884–893.

[41] Fink, B., & Penton-Voak, I. (2002). Evolutionary psychology of facial attractiveness. *Current Directions in Psychological Science, 11*(5), 154–158.

[42] Foo, Y., Simmons, L., & Rhodes, G. (2017). Predictors of facial attractiveness and health in humans. *Scientific Reports, 7*(39731).

[43] Roberts, S. C., Havlicek, J., Flegr, J., Hruskova, M., Little, A. C., Jones, B. C., … & Petrie, M. (2004). Female facial attractiveness increases during the fertile phase of the menstrual cycle. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(suppl_5).

[44] Bobst, C., & Lobmaier, J. S. (2012). Men's preference for the ovulating female is triggered by subtle face shape differences. *Hormones and Behavior, 62*(4), 413-417.

[45] Itriyeva, K. (2022). The normal menstrual cycle. *Current Problems in Pediatric and Adolescent Health Care*, *52*(5), 101183.

[46] Mihm, M., Gangooly, S., & Muttukrishna, S. (2011). The normal menstrual cycle in women. *Animal Reproduction Science*, *124*(3), 229–236.

[47] Reed, B. G., & Carr, B. R. (2000). The Normal Menstrual Cycle and the Control of Ovulation. In K. R. Feingold, B. Anawalt, M. R. Blackman, A. Boyce, G. Chrousos, E. Corpas, … & D. P. Wilson (Eds.), *Endotext*. MDText.com, Inc.

[48] Fraccaro, P. J., Feinberg, D. R., DeBruine, L. M., Little, A. C., Watkins, C. D., & Jones, B. C. (2010). Correlated male preferences for femininity in female faces and voices. *Evolutionary Psychology, 8*(3), 447–461.

[49] Jones, D., Brace, C. L., Jankowiak, W., Laland, K. N., Musselmann, L. E., Langlois, J. H., & Symons, D. (1995). Sexual selection, physical attractiveness, and facial neoteny: Cross-cultural evidence and implications. *Current Anthropology, 36*(5), 723–748.

[50] Kleisner, K., Tureček, P., Saribay, S. A., Pavlovič, O., Leongómez, J. D., Roberts, S. C., … & Varella, M. A. C. (2024). Distinctiveness and femininity, rather than symmetry and masculinity, affect facial attractiveness across the world. *Evolution and Human Behavior, 45*(1), 82-90.

[51] Moore, F. R., Law Smith, M. J., Taylor, V., & Perrett, D. I. (2011). Sexual dimorphism in the female face is a cue to health and social status but not age. *Personality and Individual Differences, 50*(7), 1068-1073.

[52] Perrett, D. I., May, K. A., & Yoshikawa, S. (1994). Facial shape and judgments of female attractiveness. *Nature*, 368, 239–242.

[53] Sutić, D., Brešković, I., Huić, R., & Jukić, I. (2010). "Automatic evaluation of facial attractiveness," The 33rd International Convention MIPRO, Opatija, Croatia, 1339-1342.

[54] Zaidel, D. W., & Cohen, J. A. (2005). The face, beauty, and symmetry: Perceiving asymmetry in beautiful faces. *International Journal of Neuroscience, 115*(8), 1165-1173.

[55] Fan, J., Chau, K. P., Wan, X., Zhai, L., & Lau, E. (2012). Prediction of facial attractiveness from facial proportions. *Pattern Recognition, 45*(6), 2326-2334.

[56] Bougourzi, F., Dornaika, F., & Taleb-Ahmed, A. (2022). Deep learning based face beauty prediction via dynamic robust losses and ensemble regression. *Knowledge-Based Systems*, 242.

[57] Gan, J., Xie, X., Zhai, Y., He, G., Mai, C., & Luo, H. (2022). Facial beauty prediction fusing transfer learning and broad learning system. *Soft Computing, 27*, 13391-13404.

[58] Gao, L., Li, W., Huang, Z., Huang, D., & Wang, Y. (2018). *Automatic facial attractiveness prediction by deep multi-task learning*. 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 3592-3597.

[59] Simonyan K., & Zisserman A. (2014). Very deep convolutional networks for large-scale image recognition. *Computer Vision and Pattern Recognition*.

[60] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *Computer Vision and Pattern Recognition*.

[61] Gan, J., Xiang, L., Zhai, Y., Chaoyun, M., He, G., Zeng, J., … & Piuri, V.-C. (2020). 2M BeautyNet: Facial beauty prediction based on multi-task transfer learning. *IEEE Access, 8*, 20245-20256.

[62] Vahdati, E., & Suen, C. Y. (2019). Female Facial Beauty Analysis Using Transfer Learning and Stacking Ensemble Model. In: Karray, F., Campilho, A., Yu, A. (eds) Image Analysis and Recognition. ICIAR 2019. Lecture Notes in Computer Science, 11663. Springer, Cham.

[63] Zhai, Y., Yu, H., Ying, X., Junying, G., He, C., Wenbo, D., … & Fabio, S. (2020). Asian Female Facial Beauty Prediction Using Deep Neural Networks via Transfer Learning and Multi-Channel Feature Fusion. *IEEE Access, 8*, 56892-56907.

[64] Liang, L., Lin, L., Jin, L., Xie, D., & Li, M. (2018). SCUT-FBP5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction.

[65] Fink, B. Neave, N., Manning, J. T., & Grammer. K. (2006). Facial symmetry and judgements of attractiveness, health and personality. *Personality and Individual Differences, 41*(3), 491-499.

[66] Jones, B. C., Little, A. C., Penton-Voak, I. S., Tiddeman, B. P., Burt, D. M., & Perrett, D. I. (2001). Facial symmetry and judgements of apparent health: Support for a "good genes" explanation of the attractiveness–symmetry relationship. *Evolution and Human Behavior, 22*(6), 417-429.

[67] Perrett, D. I., Burt, D. M., Penton-Voak, I. S., Lee, K. J., Rowland, D. A., & Edwards, R. (1999). Symmetry and human facial attractiveness. *Evolution and Human Behavior, 20*(5), 295-307.

[68] Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*(4), 1122-1135.

[69] Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces—KDEF (CD ROM). Stockholm: Karolinska Institute, Department of Clinical Neuroscience, Psychology Section.

[70] Ebner, N., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior research Methods, 42*(1), 351-362.

[71] DeBruine, L., & Benedict, J. (2017): Face Research Lab London Set. figshare. Dataset.

[72] Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 6*(2), 107-116.

[73] Khan, S., Rahmani, H., Shah, S. A. A., & Bennamoun, M. (2018). CNN Learning. In: A Guide to Convolutional Neural Networks for Computer Vision. Synthesis Lectures on Computer Vision. Springer, Cham.

[74] Garrido, M. V., & Prada, M. (2017). KDEF-PT: Valence, emotional intensity, familiarity and attractiveness ratings of angry, neutral, and happy faces. *Frontiers in Psychology, 8*(2181).

[75] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2016). Inception-v4, Inception-ResNet and the impact of residual connections on learning. *Computer Vision and Pattern Recognition.*

[76] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., … & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *Computer Vision and Pattern Recognition*.

[77] Tan, M., & Le, Q. V. (2021). EfficientNetV2: Smaller models and faster training. *Computer Vision and Pattern Recognition*.

[78] Chollet, F. (2016). Xception: Deep Learning with Depthwise Separable Convolutions. *Computer Vision and Pattern Recognition.*

[79] Chollet, F. (2015). Keras. GitHub repository. https://github.com/fchollet/keras

[80] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *Machine Learning*.

[81] Li, H. Chaudhari, P., Yang, H., Lam, M., Ravichandran, A., Bhotika, R., & Soatto, S. (2020). Rethinking the Hyperparameters for Fine-tuning. *Computer Vision and Pattern Recognition*.

[82] Tiddeman, B., & Perrett, D. (2002). Transformation of dynamic facial image sequences using static 2D prototypes. *The Visual Computer, 18*(4), 218-225.

[83] DeBruine, L. (2018). debruine/webmorph: Beta release 2 (v0.0.0.9001). Zenodo.

[84] DeBruine, L. M., Jones, B. C., Little, A. C., Boothroyd, L. G., Perrett, D. I., Penton-Voak, I. S., … & Tiddeman, B. P. (2006). Correlated preferences for facial masculinity and ideal or actual partner's masculinity. *Proceedings of the Royal Society B: Biological Sciences*, 273, 1355–1360.

[85] Penton-Voak, I. S., Little, A. C., Jones, B. C., Burt, D. M., Tiddeman, B. P., & Perrett, D. I. (2003). Female condition influences preferences for sexual dimorphism in faces of male humans (*Homo sapiens*). *Journal of Comparative Psychology, 117*(3), 264–271.

[86] Rowland, D. A., & Perrett, D. I. (1995). Manipulating facial appearance through shape and color. *IEEE computer graphics and applications*, *15*(5), 70-76.

[87] Gangestad, S. W., & Thornhill, R. (2003). Facial masculinity and fluctuating asymmetry. *Evolution and Human Behavior, 24*(4), 231–241.

[88] Pinheiro, J., & Bates, D. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer.

[89] R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[90] Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

[91] Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.

[92] Fernandes, N. F. R. (2024). Analyzing VGG-19's Bias in Facial Beauty Prediction: Preference for Feminine Features. https://doi.org/10.17605/OSF.IO/RW4D3

[93] Marcinkowska, U. M, Kozlov, M. V., Cai, H., Contreras-Garduño, J., Dixson, B. J., Oana, G. A., … & Rantala, M. J. (2014). Cross-cultural variation in men's preference for sexual dimorphism in women's faces. *Biology Letters, 10*(4), 20130850.

[94] O'Connor, J. J., Fraccaro, P. J., Pisanski, K., Tigue, C. C., & Feinberg, D. R. (2013). Men's preferences for women's femininity in dynamic cross-modal stimuli. *PLoS One, 8*(7), e69531.

[95] Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour, 69*(3), 561-568.

[96] Johnston, V. S., Hagel, R., Franklin, M., Fink, B., & Grammer, K. (2001). Male facial attractiveness: evidence for hormone-mediated adaptive design. *Evolution and Human Behavior, 22*(4), 251-267.

[97] Little, A. C., Jones, B. C., Waitt, C., Tiddeman, B. P., Feinberg, D. R., Perrett, D. I., … & Reimchen, T. (2008). Symmetry is related to sexual dimorphism in faces: data across culture and species. *PLoS ONE 3*(5), e2106.

[98] Penton-Voak, I. S., Perrett, D. I., Castles, D. L., Kobayashi, T., Burt, D. M., Murray, L. K., & Minamisawa, R. (1999). Menstrual cycle alters face preference. *Nature, 399*, 741–742.

[99] Penton-Voak, I. S., & Perrett, D. I. (2001). Male facial attractiveness: Perceived personality and shifting female preferences for male traits across the menstrual cycle. *Advances in the Study of Behavior, 30,* 219 – 260.

[100] Stower, R. E., Lee, A. J., McIntosh, T. L., Sidari, M. J., Sherlock, J. M., & Dixson, B. J. (2020). Mating strategies and the masculinity paradox: how relationship context, relationship status, and sociosexuality shape women's preferences for facial masculinity and beardedness. *Archives of Sexual Behavior,* 49, 809–820.

[101] DeBruine, L. M., Jones, B. C., Smith, F. G. & Little, A. C. (2010). Are attractive men's faces masculine or feminine? The importance of controlling confounds in face stimuli. *Journal of Experimental Psychology. Human Perception and Performance, 36*(3), 751–758.

[102] Little, A. C., Burt, D. M., Penton-Voak, I. S., & Perrett, D. I. (2001). Self-perceived attractiveness influences human female preferences for sexual dimorphism and symmetry in male faces. *Proceedings of the Royal Society of London, Series B, 268*(1462), 39–44.

[103] Marcinkowska, U. M., Rantala, M. J., Lee, A. J.. Kozlov, M. V., Aavik, T., Cai, H., ... & Dixson, B. J. W. (2019). Women's preferences

for men's facial masculinity are strongest under favorable ecological conditions. *Scientific Reports,* 9(3387).

[104] Nakamura, K., & Watanabe, K. (2020). A new data-driven mathematical model dissociates attractiveness from sexual dimorphism of human faces. *Scientific Reports, 10*(16588).

[105] Cornwell, R. E., Boothroyd, L., Burt, D. M., Feinberg, D. R., Jones, B. C., Little, A. C., … & Perrett, D. I. (2004). Concordant preferences for opposite-sex signals? Human pheromones and facial characteristics. *Proceedings of the Royal Society of London, Series B, 271*, 635–640.

[106] Swaddle, J. P., & Riersen, G. W. (2002). Testosterone increases perceived dominance but not attractiveness in human males. *Proceedings of the Royal Society of London, Series B, 269*(1507), 2285–2289.

[107] Hu, Y., Abbasi, N. H., Zhang, Y., & Chen, H. (2018). The effect of target sex, sexual dimorphism, and facial attractiveness on perceptions of target attractiveness and trustworthiness. *Frontiers in Psychology, 9*.

[108] Welling, L. L. M., Jones, B. C., DeBruine, L. M., Little, A. C., & Smith, F. G. (2008). Exposure to sexually attractive men decreases women's preferences for feminine faces. *Journal of Evolutionary Psychology, 6*(3), 219–230.

[109] Little, A. C., Burriss, R. P., Jones, B. C., DeBruine, L. M., & Caldwell, C. A. (2008). Social influence in human face preference: men and women are influenced more for long-term than short-term attractiveness decisions. *Evolution and Human Behavior, 29*(2), 140-146,

[110] Bartlome, R. I., & Lee, A. J. (2023). Facial attractiveness, but not facial masculinity, is used as a cue to paternal involvement in fathers. *Adaptive Human Behavior and Physiology,* 9, 182–197.

[111] Jones, B. C., DeBruine, L. M., Perrett, D. I., Little, A. C., Feinberg, D. R., & Law Smith, M. J. (2008). Effects of menstrual cycle phase on face preferences. *Archives of Sexual Behavior*, *37*(1), 78–84.

[112] Fink, B., & Neave, N. (2005). The biology of facial beauty. *International Journal Cosmetic Science, 27*(6), *317–325.*

[113] Schölkopf, B., Platt, J., & Hofmann, T. (2007). *A Humanlike predictor of facial attractiveness*. Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference, MIT Press, 649-656.

[114] Tybur, J. M., Fan, L., Jones, B. C., Holzleitner, I. J., Lee, A. J., & DeBruine, L. M. (2022). Re-evaluating the relationship between pathogen avoidance and preferences for facial symmetry and sexual dimorphism: A registered report. *Evolution and Human Behavior, 43*(3), 212-223.

[115] Scheib, J. E., Gangestad, S. W., & Thornhill, R. (1999). Facial attractiveness, symmetry, and cues to good genes. *Proceedings of the Royal Society of London, B, 266,* 1913–1917.

[116] Penton-Voak, I. S., & Chen, J. Y. (2004). High salivary testosterone is linked to masculine male facial appearance in humans. *Evolution and Human Behavior, 25*(4), 229–241.

[117] Mogilski, J. K., & Welling, L. L. M. (2017). The relative importance of sexual dimorphism, fluctuating asymmetry, and color cues to health during evaluation of potential partners' facial photographs. *Human Nature, 28*, 53–75.

[118] Swaddle, J. P., & Cuthill, I. C. (1995). Asymmetry and human facial attractiveness: symmetry may not always be beautiful. *Proceedings of the Royal Society B, 261*, 111–116.

[119] Zheng, R., Ren, D., Xie, C., Pan, J., & Zhou, G. (2021). Normality mediates the effect of symmetry on facial attractiveness. *Acta Psychologica, 217*.

[120] Palumbo, R., Adams, R. B., Hess, U., Kleck, R. E., & Zebrowitz, L. (2017). Age and Gender Differences in Facial Attractiveness, but Not Emotion Resemblance, Contribute to Age and Gender Stereotypes. *Frontiers in Psychology, 8*.

[121] McLellan, B., & McKelvie, S. J. (1993). Effects of age and gender on perceived facial attractiveness. *Canadian Journal of Behavioural Science, 25*(1), 135–142.

[122] Levy, B., Ariely D., Mazar, N., Chi, W., Lukas, S., & Elman, I. (2008). Gender differences in the motivational processing of facial beauty. *Learning and Motivation, 39*(2), 136–145.

[123] Nistor, S. C., Marina, A. -C., Darabant, A. S., & Borza, D. (2017). *Automatic gender recognition for "in the wild" facial images using convolutional neural networks*. 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 2017, pp. 287-291.

[124] Nga, C. H., Nguyen, K. -T., Tran, N. C., & Wang, J. -C. (2020). *Transfer learning for gender and age prediction*. 2020 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan), Taoyuan, Taiwan, pp. 1-2.

[125] Żejmo, A., Gielert, M., Grabski, M., & Kostek, B. (2023). *Assessing the attractiveness of human face based on machine learning*. *Procedia Computer Science, 225*, 1019-1027.

[126] Golle, J., Mast, F. W., & Lobmaier, J. S. (2014). Something to smile about: The interrelationship between attractiveness and emotional expression. *Cognition and Emotion, 28*(2), 298-310.

[127] Lindeberg, S., Craig, B. M., & Lipp, O. V. (2019). You look pretty happy: Attractiveness moderates emotion perception. *Emotion, 19*(6), 1070–1080.

[128] Balasubramanian, B., Diwan, P., Nadar, R., & Bhatia, A. (2019). *Analysis of Facial Emotion Recognition*. *3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, 945-949.

[129] Lim, H.-E. (2008). The use of different happiness rating scales: Bias and comparison problem?. *Social Indicators Research, 87*(2), 259–267.

[130] Allen, I. E., & Seaman, C. A. (2007). Likert Scales and Data Analyses. *Quality Progress, 40*(7), 64-65.

[131] Simpson, J. A., Gangestad, S. W., & Lerma, M. (1990). Perception of physical attractiveness: Mechanisms involved in the maintenance of romantic relationships. *Journal of Personality and Social Psychology, 59*(6), 1192–1201. https://doi.org/10.1037/0022-3514.59.6.1192

[132] Kandrik, M., & DeBruine, L. M. (2012). Self-rated attractiveness predicts preferences for opposite-sex faces, while self-rated sex-typicality predicts preferences for same-sex faces. *Journal of Evolutionary Psychology, 10*(4), 177–186.

[133] Glassenberg, A. N., Feinberg, D. R., Jones, B. C., Little A. C., & Debruine L. M. (2010). Sex-dimorphic face shape preference in heterosexual and homosexual men and women. *Archives of Sexual Behavior, 39*(6), 1289-1296.

[134] Zhang, J., Zheng, L., & Zheng, Y. (2018). Consistency in preferences for masculinity in faces, bodies, voices, and personality characteristics among homosexual men in China. *Personality and Individual Differences, 134*, 137-142.

[135] Zheng, L. (2021). The dyadic effects of top/bottom sexual self-labels and partner sexual role requirements on facial masculinity preferences among gay and bisexual men in china. *Journal of Sex Research, 58*(1), 122-12.

Real-Time Self-Localization and Mapping for Autonomous Navigation of Mobile Robots in Unknown Environments

Serik Tolenov, Batyrkhan Omarov
Joldasbekov Institute of Mechanics and Engineering, Almaty, Kazakhstan

***Abstract*—This paper delves into the progressive design and operational capabilities of advanced robotic platforms, highlighting their adaptability, precision, and utility in diverse industrial settings. Anchored by a robust modular design, these platforms integrate sophisticated sensor arrays, including LiDAR for enhanced spatial navigation, and articulated limbs for complex maneuverability, reflecting significant advancements in automation technology. We examine the architectural intricacies and technological integrations that enable these robots to perform a wide range of tasks, from material handling to intricate assembly operations. Through a detailed analysis of system configurations, we assess the implications of such technologies on efficiency and customization in automated processes. Furthermore, the paper discusses the challenges associated with the deployment of advanced robotics, including the complexities of system integration, maintenance, and the steep learning curve for operational proficiency. We also explore future directions in robotic development, emphasizing the potential integration with emerging technologies such as artificial intelligence, the Internet of Things, and augmented reality, which promise to elevate autonomous decision-making and improve human-robot interaction. This comprehensive review aims to provide insights into the current capabilities and future prospects of robotic systems, offering a perspective on how ongoing innovations may reshape industrial practices, enhance operational efficiency, and redefine the landscape of automation technology.**

Keywords*—Robotic platforms; automation technology; LiDAR navigation; system integration; artificial intelligence; Internet of Things; human-robot interaction

I. Introduction

Simultaneous Localization and Mapping (SLAM) remains a cornerstone technology in the field of robotics, facilitating the autonomous navigation of mobile robots in environments unknown and unmapped. As the demand for autonomous systems spans industries—from automotive to agricultural and healthcare—SLAM has evolved from a theoretical concept to a critical component in real-world applications. This technology enables a robot to build a map of an unknown environment while simultaneously determining its location within that map. The iterative nature of SLAM—constantly updating and refining the map and the robot's location—makes it one of the most complex yet transformative technologies in modern robotics.

SLAM's significance is particularly pronounced in environments that are dynamic and unstructured, where pre-existing maps are unavailable or insufficiently detailed. In such scenarios, robots must rely on their onboard sensors and processing capabilities to navigate effectively and perform tasks autonomously [1]. The dual challenges of localization (the robot's ability to know its position relative to the environment) and mapping (the process of constructing an accurate, real-time map of the environment) are intertwined tasks that must be solved concurrently. This is because the accuracy of localization directly impacts the quality of the map generated, and vice versa [2]. Fig. 1 demonstrates the simultaneously localization and mapping problem.

Historically, the development of SLAM has been driven by advancements in sensor technology and computational methods. Essential to the process are various sensors that provide the raw data needed for map construction and localization. These typically include LiDAR (Light Detection and Ranging), RGB-D cameras (which capture color (RGB) and depth (D) data), inertial measurement units (IMUs), and stereo cameras [3]. Each sensor type offers different advantages and constraints; for example, LiDAR sensors are highly effective in distance measurement but can be costly and complex, whereas RGB-D cameras provide rich visual and depth data but may struggle in poor lighting conditions [4].

The complexity of SLAM algorithms varies significantly depending on the specific application and the environment. Algorithms must efficiently process vast amounts of data from multiple sensors to produce accurate and reliable localization and mapping outcomes. These algorithms are generally categorized into two types: filter-based and graph-based SLAM. Filter-based methods, such as the Extended Kalman Filter (EKF) SLAM, iteratively estimate the state of the map and the robot’s location [5]. Graph-based SLAM, on the other hand, constructs a graph where the nodes represent robot poses at different times, and the edges represent spatial constraints between these poses, solved optimally to reconstruct the robot’s trajectory and the map [6].

Calibration plays a crucial role in ensuring the accuracy of the data collected by the sensors. Incorrect calibration can lead to significant errors in localization and mapping. The calibration process involves determining the intrinsic (internal characteristics) and extrinsic (spatial relationships among different sensors) parameters. Noise is another critical factor impacting the SLAM system; it encompasses any error that deviates from the true measurements, such as sensor inaccuracies or environmental factors like electromagnetic interference [7].

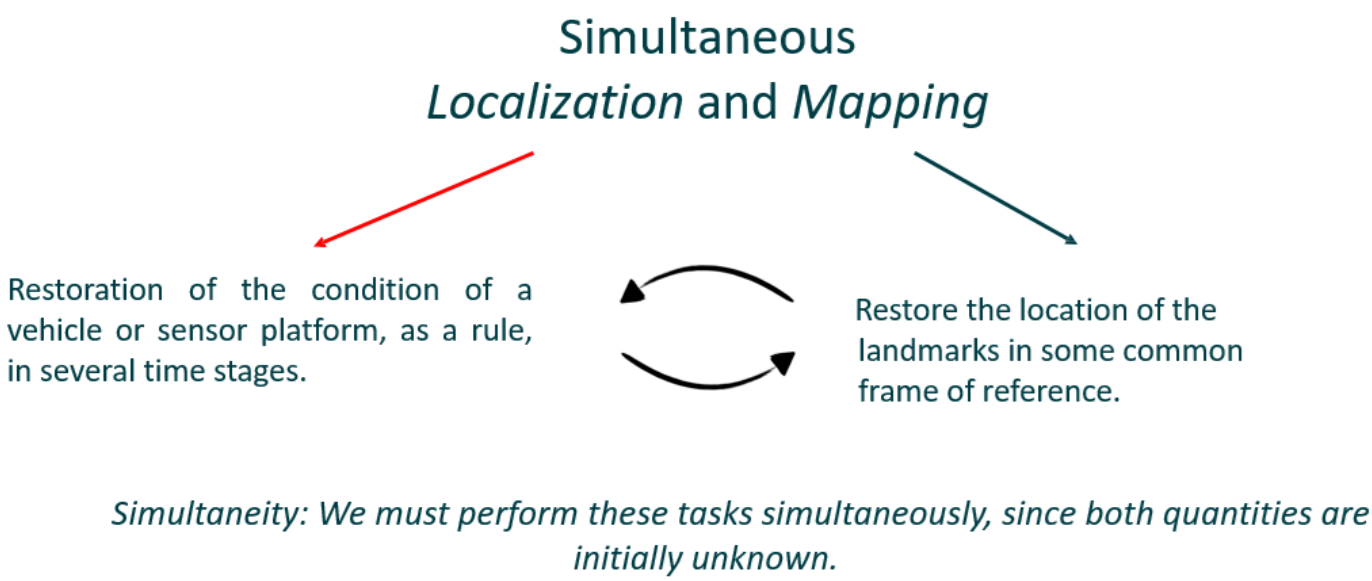


Fig. 1. Simultaneously locatiozation and mapping.

The uncertainty inherent in sensor data and the need for real-time processing make SLAM a computationally demanding task. The presence of noise and the potential for significant variability in environmental conditions mean that SLAM systems must be robust to a range of operational scenarios. As such, SLAM technology not only requires sophisticated algorithms but also powerful computational resources to handle real-time data assimilation and processing [8].

Recent developments in SLAM have seen the integration of machine learning techniques to enhance the adaptability and efficiency of SLAM systems. These approaches leverage the power of neural networks to improve feature extraction, data association, and even to predict environmental changes that might affect navigation [9]. Such innovations have opened new avenues for the application of SLAM in more complex and dynamic environments, pushing the boundaries of what autonomous robots can achieve [10].

Moreover, the integration of SLAM into various application domains has necessitated adaptations to meet specific operational requirements. For instance, autonomous vehicles use SLAM for real-time navigation and obstacle avoidance in urban environments, while agricultural robots use it to navigate between crops and perform tasks such as harvesting or planting. Each application presents unique challenges and requirements, influencing the choice of sensors, the design of algorithms, and the overall architecture of the SLAM system [11].

Overall, SLAM continues to be a vibrant field of research and application, driving forward the capabilities of autonomous robots. The ongoing evolution of sensor technologies, coupled with advances in computational algorithms and machine learning, promises to enhance the robustness, accuracy, and efficiency of SLAM systems, heralding new possibilities for automation across various sectors [12]. As this technology progresses, it will play a pivotal role in the realization of fully autonomous systems capable of operating in complex and evolving environments.

II. Problem Statement

The evolution of Simultaneous Localization and Mapping (SLAM) technology has been significantly influenced by various research efforts aimed at improving its accuracy, efficiency, and utility in diverse application domains. This section reviews the seminal and recent works that have contributed to the development of SLAM, highlighting the advancements in sensor technology, algorithmic approaches, and system integrations.

One of the foundational aspects of SLAM is the use of sophisticated sensor systems to capture environmental data essential for mapping and localization. Initial studies in SLAM primarily utilized laser range finders and sonar sensors due to their reliability in distance measurement [13]. However, with technological advancements, the use of RGB-D cameras and LiDAR sensors has become prevalent, especially in applications requiring detailed 3D mapping and object recognition [14]. These sensors not only provide depth information but also rich visual data, which is crucial for feature-based SLAM algorithms.

The development of algorithms for SLAM has seen a significant transformation from early filter-based methods to modern graph-based and machine learning-enhanced techniques. Filter-based methods, such as the Extended Kalman Filter (EKF) and the Particle Filter, have been widely used due to their robustness in online state estimation and their ability to handle the nonlinearities typical in real-world environments [15]. However, these methods often suffer from scalability issues when the environment size or the number of landmarks increases significantly [16]. This has led to the adoption of graph-based SLAM, which offers better scalability and accuracy by optimizing a graph structure that represents the spatial relationships among various poses and landmarks [17].

Graph-based SLAM algorithms, particularly those employing pose graph optimization, have revolutionized the way SLAM is implemented. These algorithms construct a network of constraints based on the relative measurements

between poses and landmarks, which is then optimized to find the most probable map configuration [18]. This approach not only improves the computational efficiency but also enhances the map's fidelity by reducing cumulative errors over long sequences.

The integration of machine learning techniques into SLAM has opened new avenues for enhancing the system's adaptability and performance. Deep learning, for instance, has been employed to improve feature extraction and association, a critical aspect of SLAM that affects the system's overall robustness and accuracy [19]. Neural networks have also been used to predict and compensate for the environmental dynamics that traditional SLAM systems might not handle well [20]. Moreover, machine learning algorithms have facilitated the development of semantic SLAM, which not only maps the environment but also understands and categorizes it into meaningful entities [21].

The calibration of sensors remains a critical issue in SLAM, as inaccurate sensor models can lead to significant localization errors. Research has focused on developing more robust calibration techniques that can be performed easily and reliably in-field [22]. These techniques ensure that the intrinsic and extrinsic parameters of the sensors are accurately determined, thus enhancing the overall reliability of the SLAM system.

Noise and uncertainty in sensor data are inherent challenges in SLAM that degrade the quality of localization and mapping. To address these issues, advanced statistical methods have been developed to model and mitigate the impact of noise and uncertainty [23]. These methods include robust estimation techniques that can identify and reject outliers in sensor data, thereby improving the SLAM system's resilience to environmental noise and sensor faults [24].

Recent advancements have also explored the use of SLAM in dynamic and unstructured environments, such as underwater, aerial, or disaster-stricken areas, where traditional SLAM techniques face significant challenges [25]. These environments require highly robust and adaptive SLAM solutions that can handle large variations in environmental conditions and sensor disruptions.

Furthermore, the application of SLAM technology has extended beyond robotic navigation to include tasks such as augmented reality, where real-time mapping and localization are crucial for overlaying virtual objects onto the physical world [26]. This demonstrates the versatility and widespread applicability of SLAM, making it a critical technology in various fields.

The related work in SLAM demonstrates a trend towards more integrated and intelligent systems capable of operating in increasingly complex environments. The continuous improvements in sensor technology, alongside innovations in computational algorithms and machine learning, are driving the evolution of SLAM towards systems that can not only navigate and map with high accuracy but also understand and interact with their environments in sophisticated ways [27-29].

Overall, the body of work surrounding SLAM encompasses a broad spectrum of research areas, including sensor technologies, algorithmic strategies, system design, and applications. Each study contributes to building a more comprehensive understanding of how autonomous systems can effectively perceive and navigate our world. The ongoing research and developments promise to enhance the capabilities of SLAM, pushing the boundaries of what autonomous systems can achieve, and broadening the horizon for future applications.

III. CHALLENGES

The primary challenge addressed in this research is the development of a robust Simultaneous Localization and Mapping (SLAM) system that efficiently integrates various sensory inputs to accurately navigate and map an environment in real-time. This entails the construction and refinement of a dynamic model that not only interprets and assimilates data from multiple sensors but also accounts for the inherent uncertainties and potential errors in sensor outputs. The effectiveness of a SLAM system hinges on its ability to synthesize this data to produce reliable, real-time updates of both the system's location (localization) and the structure of the environment (mapping).

Fig. 2. Sensor setup.

Fig. 2 displays various sensors commonly used in SLAM systems, including Cameras, Inertial Measurement Units (IMUs), LiDAR/Rangefinders, and RGB-D/Structured Light sensors. Each sensor type provides unique data essential for comprehensive environmental perception. Cameras offer visual information, IMUs provide movement and orientation data, LiDARs deliver precise distance measurements, and RGB-D sensors combine depth perception with visual data [30]. The integration of these diverse data streams is crucial for the development of a detailed and accurate map of the environment, as well as for the precise localization of the SLAM system within it.

Fig. 3. Sensor integration and mapping process.

Fig. 3 illustrates the process of sensor data integration and subsequent map generation. It shows a schematic representation of an autonomous vehicle (or drone), highlighting its trajectory and the environmental mapping it performs. Key aspects such as ego-motion estimation and calibration parameters are emphasized, indicating their roles in refining the system's performance. This figure underscores the dynamic interaction

between motion data and environmental mapping, crucial for real-time navigation and obstacle avoidance in autonomous systems [31].

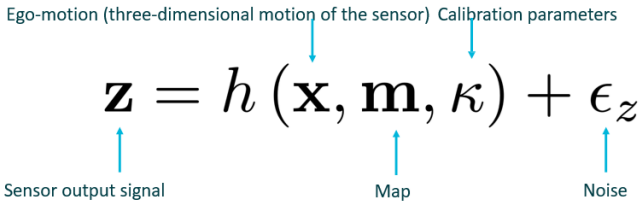


Fig. 4. SLAM system algorithmic process.

Fig. 4 introduces the mathematical model underlying the SLAM process. It describes the relationship between the sensor output signal (z), the system's state (x), the map (m), and the calibration parameters (κ), incorporating noise (εz) to account for measurement uncertainties. This model is central to the SLAM system as it forms the basis for algorithmic interpretations and adjustments made during the mapping and localization processes.

Expanding on the mathematical foundations, depicts the algorithmic workflow of a SLAM system. It illustrates how sensor inputs are transformed into a set of equations that the SLAM algorithm processes to update estimates of the system's state, the map, and calibration parameters [32]. This sequential representation highlights the continuous feedback loop essential for adaptive and responsive SLAM operations.

Also, Fig. 5 demonstrates the output from the SLAM system, showcasing the refined estimates of the system's state and the detailed environmental map. This visualization not only serves as a validation of the system's effectiveness but also illustrates the practical application of SLAM in real-world navigation scenarios.

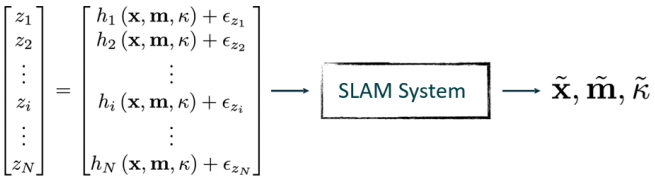


Fig. 5. Output visualization.

These figures collectively delineate the complexity of developing an effective SLAM system. They highlight the integration of multi-sensor data, the importance of accurate mathematical modeling, and the necessity of adaptive algorithms capable of continuous learning and refinement. Addressing these challenges through innovative solutions is fundamental to advancing the field of autonomous navigation and ensuring the practical deployment of SLAM systems in diverse and dynamically changing environments.

IV. Materials and Methods

In this section, we delineate the technical specifications and procedural framework utilized in the development and testing of the robotic platform. This section is structured to provide a comprehensive overview of the equipment, software, and methodologies employed to achieve the objectives stated in the study. It begins with a detailed description of the robotic system's hardware configuration, including all sensors and actuaries involved, followed by an exposition of the software algorithms used for tasks such as navigation, mapping, and task execution. We also outline the experimental setup and the conditions under which the robot was tested, highlighting any simulations or real-world scenarios that were employed. This thorough detailing ensures transparency and reproducibility of the results, offering insights into the practical applications and limitations of the robotic system in various environments. By elucidating these methods, we aim to provide a clear path for future research and development in robotic technologies, ensuring that subsequent innovations can build upon a solid foundation of well-documented experimental practices.

Fig. 6 illustrates the impact of sensor choice on the configuration of maps generated in a SLAM system. The figure presents three distinct visual representations derived from different sensor technologies, emphasizing how each sensor type influences the nature and detail of the resulting maps.

1) Left panel: Collection of Photos: This panel displays a point cloud visualization generated from a collection of photographs, likely obtained using photogrammetry techniques. The image showcases a highly detailed three-dimensional reconstruction of a complex architectural structure. The density and granularity of the points indicate high spatial resolution, allowing for detailed feature extraction and textural information which are crucial for creating accurate visual maps.

2) Middle panel: 2D LiDAR Scan: The middle image depicts a 2D LiDAR scan, characterized by its line-based representation which outlines a simple closed trajectory within a bounded environment. This scan typically provides precise distance measurements from the sensor to the surrounding obstacles, represented here by the clear, unambiguous lines [32]. The simplicity and clarity of the data focus on spatial relationships and obstacle detection, suitable for navigation and basic mapping tasks.

3) Right panel: 3D Image: The rightmost image demonstrates a 3D map constructed using advanced imaging techniques, possibly involving a combination of LiDAR and structured light sensors. This representation is not only three-dimensional but also includes color coding and textural overlays, indicating variability in elevation and possibly the integration of additional data types like thermal or multispectral imaging [33]. Such detailed maps are instrumental in applications requiring in-depth environmental analysis and feature-rich navigation.

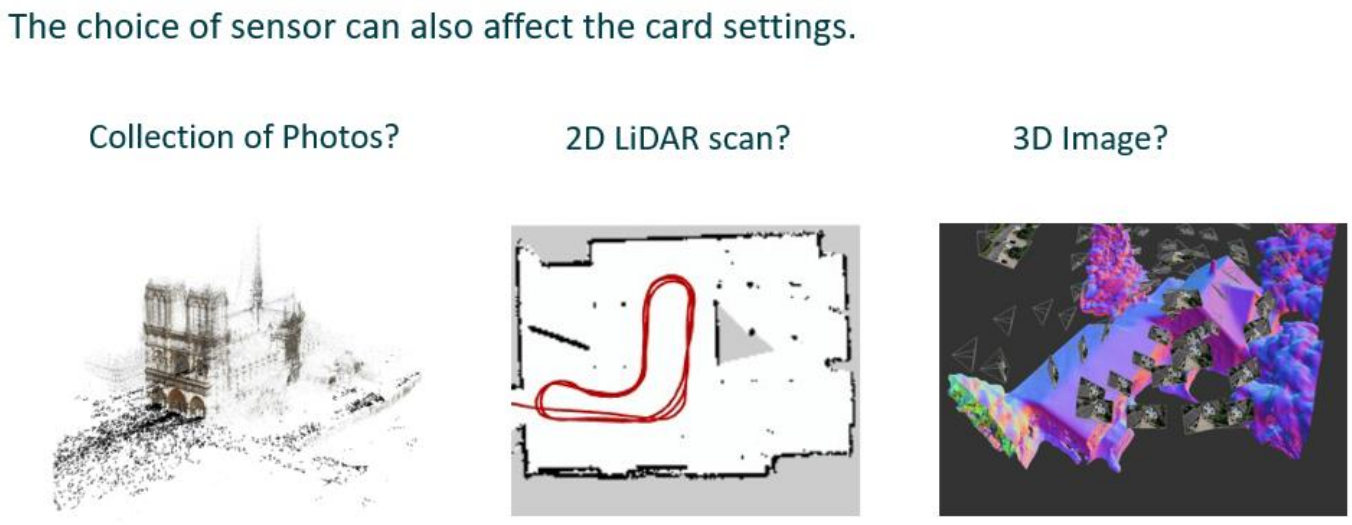


Fig. 6. The choice of sensor can also affect the card settings.

Fig. 6 presents an insightful depiction of the profound impact that various sensor technologies have on the fidelity, detail, and application suitability of maps generated through Simultaneous Localization and Mapping (SLAM) systems. This illustration serves to emphasize the pivotal role of sensor selection in the design and operational efficacy of SLAM-based applications [34], showcasing how each sensor type distinctly enhances the system's overall functionality and accuracy. The diversity in sensor capabilities is exemplified by the use of photogrammetry, which can yield highly detailed three-dimensional reconstructions; these are particularly valuable in scenarios where visual detail is critical. Conversely, 2D LiDAR scans, known for their precision and simplicity, become indispensable in navigation tasks focused on obstacle avoidance [35]. Moreover, the integration of 3D imaging techniques—which merge depth perception with high-resolution imagery—affords comprehensive environmental mapping capabilities. These are crucial for executing more intricate navigation and interaction tasks in robotic applications, facilitating advanced maneuvers and interactions within complex environments [36]. The strategic selection of sensors, therefore, not only influences the quality of the data collected but also significantly affects the adaptability and applicability of the SLAM system to varied operational contexts. As such, Fig. 6 not only highlights the technical considerations in sensor choice but also underscores their strategic importance in enhancing the robustness and versatility of SLAM technologies in dynamic settings.

The choice of sensor not only affects the quality of the data collected but also dictates the SLAM system's ability to function under various environmental conditions. For example, photogrammetric methods might be less effective in poorly lit conditions, whereas LiDAR systems can operate effectively in a range of lighting scenarios, offering versatility across different operational settings [37]. This adaptability is crucial for tailoring the SLAM approach to meet specific requirements, whether in indoor environments, outdoor landscapes, or areas with variable lighting and weather conditions.

Fig. 6 vividly captures the essence of how sensor technology choices directly influence the design and efficiency of SLAM systems. By highlighting the differences in map fidelity and application suitability offered by various sensors, the figure emphasizes the importance of strategic sensor selection. This ensures that SLAM systems are not only optimized for specific tasks but are also capable of performing effectively under the unique constraints of each application environment.

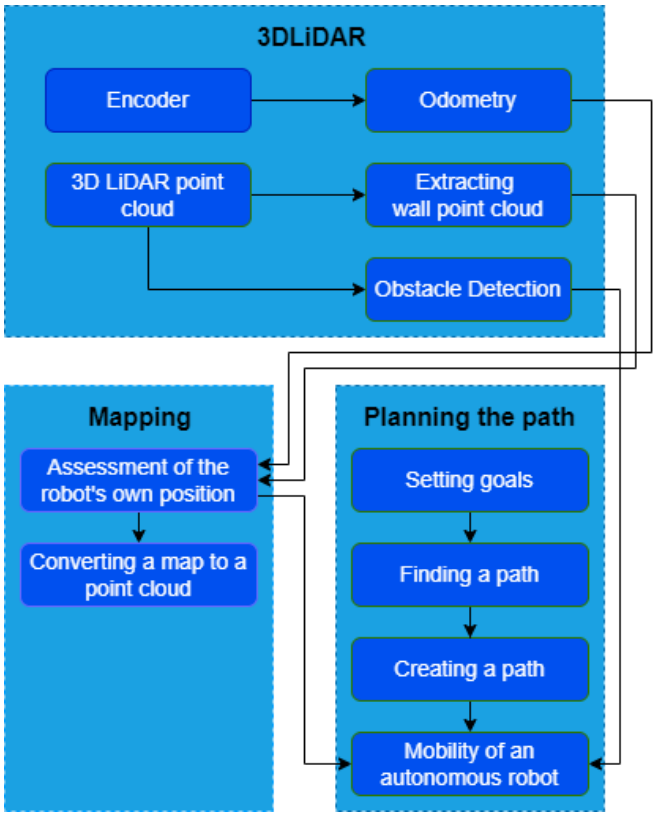


Fig. 7. Flowchart of the proposed system for navigation of the mobile robots.

Fig. 7 provides a schematic representation of the workflow involved in using 3D LiDAR technology for Simultaneous Localization and Mapping (SLAM) and path planning in autonomous robots. The diagram is divided into two main sections—Mapping and Planning the path—each containing multiple subprocesses that detail the steps and components involved in SLAM operations using 3D LiDAR.

1) 3D LiDAR Module:

a) Encoder: This component is responsible for encoding the raw data collected from the LiDAR sensor, preparing it for further processing.

b) Odometry: It computes the robot's change in position over time by analyzing the sequential data points captured by the LiDAR, which is crucial for tracking the robot's trajectory.

c) 3D LiDAR point cloud: This is the raw output from the LiDAR sensor, which includes a three-dimensional set of data points representing the distances to the surrounding objects.

d) Extracting wall point cloud: This process involves filtering the 3D point cloud to identify points that correspond to wall surfaces, which are critical for defining the structure of the environment.

e) Obstacle detection: Utilizing the data from the point cloud, this function detects and localizes objects that could impede the navigation of the robot, ensuring safe movement within the environment.

2) Mapping:

a) Assessment of the robot's own position: This step integrates the odometry and sensor data to determine the robot's current position within the map being constructed.

b) Converting a map to a point cloud: In this final step of the mapping phase, the map generated by the robot is converted back into a point cloud format, which can be used for further refinement and verification of the map's accuracy.

3) Planning the path

a) Setting goals: This initial step involves defining the destination or specific waypoints that the robot needs to reach.

b) Finding a path: Algorithms compute the optimal path to the goal, considering the map data and any detected obstacles.

c) Creating a path: This involves generating a navigable path, which includes adjusting for dynamic changes in the environment that might affect the planned route.

d) Mobility of an autonomous robot: The final output is the execution of the path, allowing the robot to move autonomously towards its goal while dynamically adjusting its course as necessary.

This flowchart effectively encapsulates the complex interactions between different components of a SLAM system equipped with 3D LiDAR technology. It highlights the critical role of each subprocess in achieving accurate localization, comprehensive environmental mapping, and efficient path planning, which are essential for the autonomy of mobile robots. The clarity and structure of the diagram provide a clear understanding of how these technologies are integrated to support the navigation capabilities of robotic systems.

V. Experimental Results

Fig. 8 showcases a graphical depiction of a 2D SLAM process visualized through a Python programming interface. The display captures a real-time simulation where the SLAM algorithm maps an environment based on data collected from a mobile robot equipped with LiDAR sensors. The predominant red lines illustrate the robot's orientation and the trajectory as it navigates the simulated setting. These lines represent both the paths traversed by the robot and the physical boundaries encountered, such as walls and obstacles, detected through its sensor array.

The demonstration environment is meticulously configured within a Python integrated development environment (IDE), underscoring the significant reliance on Python for the development and execution of the SLAM (Simultaneous Localization and Mapping) algorithm. This utilization of Python is emblematic of contemporary practices in the fields of robotics and automation, where Python's versatility and robust library support facilitate complex algorithmic development and testing [38]. The interface presents a dynamic visualization of the robot's movement, illustrating the ongoing process of mapping and localization that is central to SLAM operations. As the robot navigates through its environment, the visualization updates to display a network of red lines, representing the robot's path and the environmental boundaries it detects.

These red lines are not static; they evolve continuously as the robot acquires and processes new sensor data, thereby adjusting its internal map of the surroundings in real-time. This feature highlights the adaptive nature of SLAM systems, which must constantly refine their calculations based on incoming data to maintain accurate navigation and mapping. The ability to visually track these adjustments in real-time provides invaluable feedback during development and testing, allowing for immediate identification and resolution of any discrepancies in the algorithm's performance. Overall, the use of the Python IDE in this context not only enhances the efficiency of algorithm development but also enriches the analytical capabilities essential for advancing SLAM technology in robotic applications.

On the left side of the screen, a portion of the Python code involved in the SLAM operation is visible, hinting at the initialization of environment parameters and the integration of sensor data. In the bottom left corner, command line outputs possibly display log messages or computational diagnostics, aiding in the debugging and optimization of the algorithm. This visual and technical presentation in Fig. 8 not only underscores the practical application of SLAM technology in simulated environments for educational and development purposes but also highlights the robust capabilities of autonomous navigation systems in continuously updating and refining their understanding of complex environments.

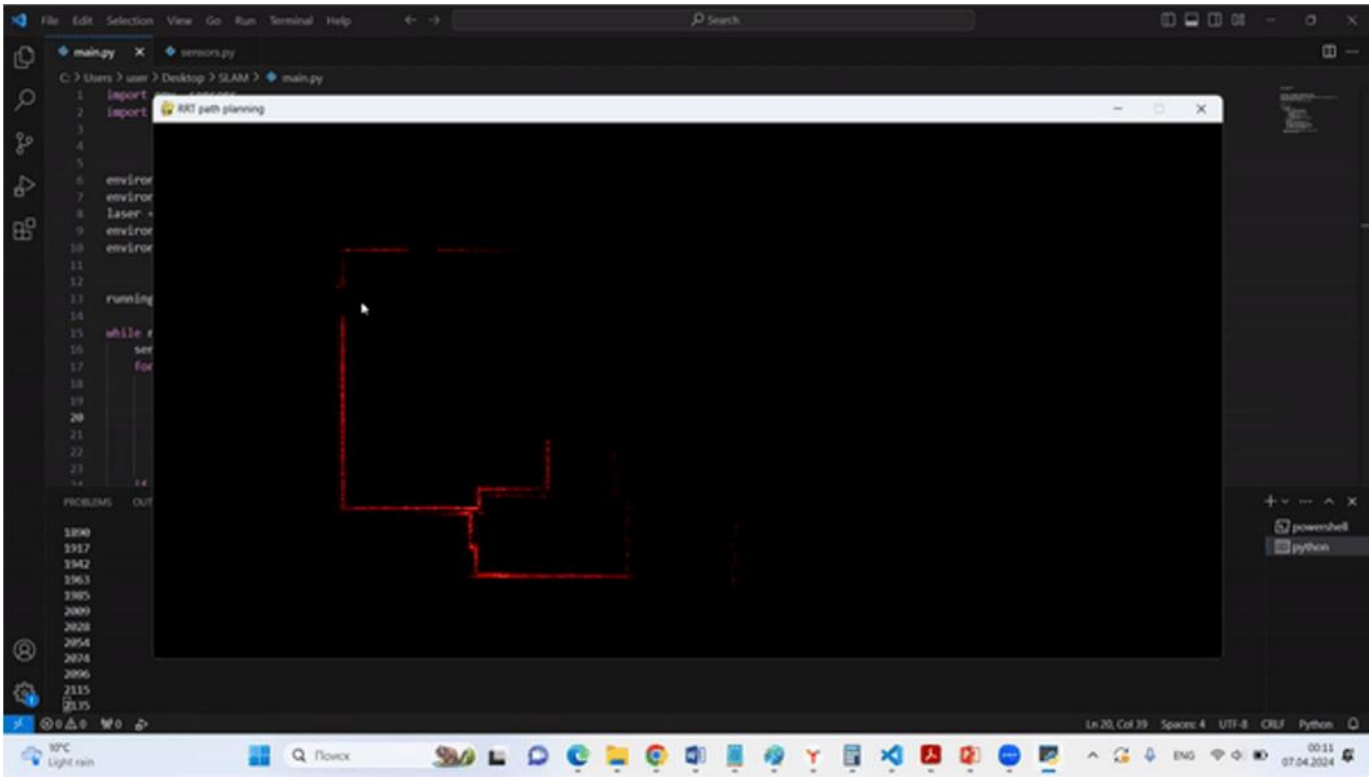


Fig. 8. Simultaneous start of localization and mapping.

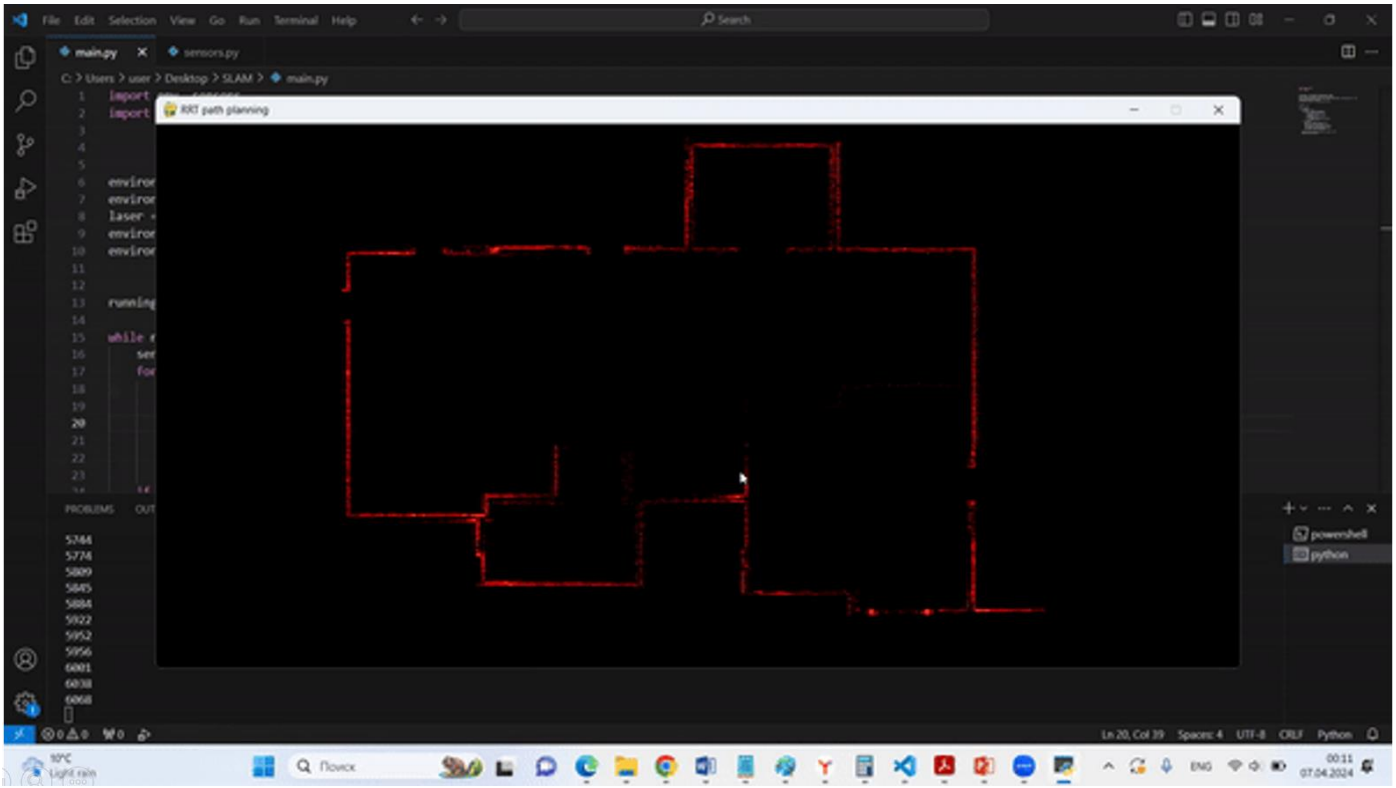


Fig. 9. Simultaneous localization and mapping for all rooms.

Fig. 9 provides a vivid depiction of a 2D Simultaneous Localization and Mapping (SLAM) process, visualized within a Python programming environment on a desktop computer's display. This image captures a real-time view of a SLAM algorithm in action, dynamically mapping an environment as navigated by a mobile robot. The visualization is characterized

by prominent red lines that illustrate the robot's trajectory and the environmental boundaries it detects. These lines, generated from data acquired via LiDAR sensors equipped on the robot, effectively demonstrate the robot's capacity to measure distances to various obstacles such as walls and other physical barriers. This functionality highlights the sophisticated integration of advanced sensor technology, which empowers the robot to navigate and map its surroundings with exceptional precision [39]. The display not only serves as a practical interface for observing the SLAM process but also underscores the critical role of advanced sensing and computational technologies in enhancing robotic navigation and environmental interaction.

Technically, this figure displays an open Python script on the left side of the computer screen, signifying that the SLAM process is governed by custom-developed software. This script is replete with various functions and libraries crucial for running the SLAM algorithm, incorporating components that manage sensor inputs and undertake intricate computational geometry calculations. The right side of the screen vividly exhibits the mapping output, which is dynamically updated, reflecting the environmental model as the robot processes new sensor data and recalibrates its trajectory accordingly. This facet of the interface not only visualizes the map's progressive evolution in response to the robot's movements but also accentuates the algorithm's capacity to swiftly adapt and refine the navigational path in real-time. This dynamic adaptation is crucial for maintaining accurate and reliable navigation, particularly in complex or changing environments, showcasing the robustness and flexibility of the SLAM technology in practical applications.

This visualization is pivotal in demonstrating the robot's adeptness in path planning and navigation, effectively showcasing the seamless integration of robotics, programming, and advanced sensor technologies. The ongoing updates to the environmental map, dictated by the robot's movements, underline the formidable real-time processing capabilities inherent in contemporary SLAM technologies. These advancements facilitate sophisticated navigation and autonomous movement, underscoring their capacity to address complex navigational challenges. Moreover, the robot's ability to recalibrate its trajectory based on live data exemplifies the dynamic interplay between the robotic hardware and its controlling software. This interaction highlights the critical role of SLAM in enhancing robotic responsiveness and improving operational efficiency across diverse applications. The depiction not only illustrates the technical sophistication of modern robotics but also points to broader implications for the future development of autonomous systems. These systems are increasingly capable of operating in varied and dynamically changing environments, pushing the boundaries of what can be achieved with autonomous robotic technology. This progress in SLAM technology not only enhances current robotic applications but also paves the way for future innovations in the automation sector.

Fig. 10 illustrates a mobile robotic unit equipped with a LiDAR sensor, demonstrating its navigation capabilities in an indoor environment with a wooden floor and plain walls. This robot, a compact, multi-wheeled vehicle, showcases advanced autonomous maneuverability facilitated by the mounted LiDAR system, which is visible on the top of the unit. The LiDAR sensor plays a critical role in the robot's ability to perceive its surroundings, enabling it to navigate smoothly and avoid collisions with obstacles that are not present in this particular scenario.

In the displayed action, the robot initiates a turn, which appears to be part of a routine exploration or mapping activity. The motion of the robot is fluid and precise, indicating robust control mechanisms and effective real-time processing of the LiDAR data to guide its movements. The environment itself, though minimally featured, serves as an ideal testing ground to illustrate the robot's basic operational functions, such as turning and straightforward locomotion.

Fig. 10. The proposed mobile robot.

This demonstration highlights the practical application of LiDAR technology in robotics for tasks that require autonomous exploration and mapping. The robot's design suggests it is possibly suited for a variety of applications, from domestic assistance to more complex industrial tasks that require navigation in tight spaces. The integration of LiDAR with the robotic hardware exemplifies the synergy between mechanical design and sensor technology, which is pivotal in advancing the capabilities of autonomous robotic systems.

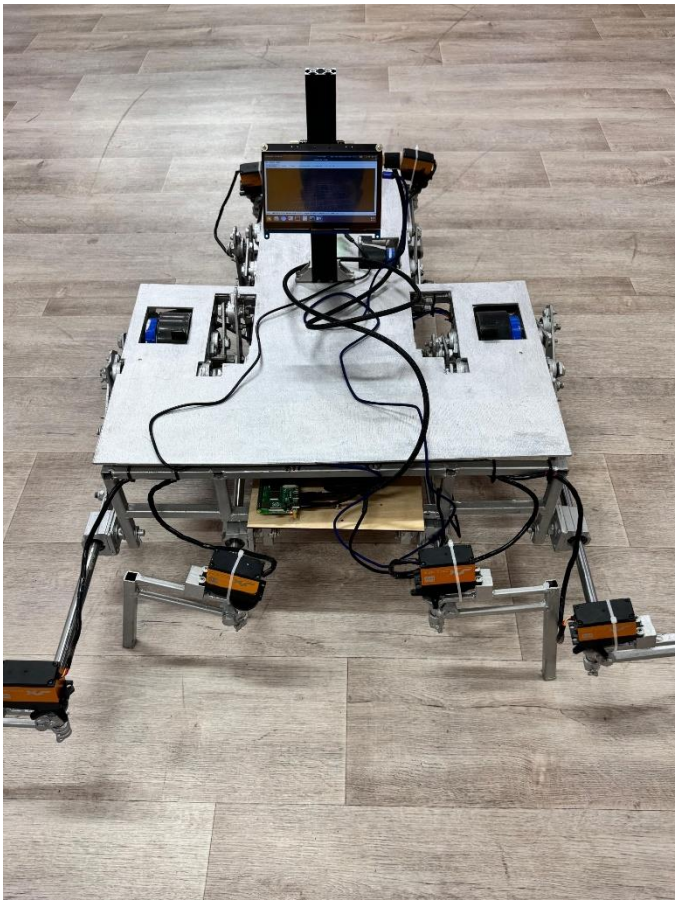


Fig. 11. A walking robot with simultaneous localization and mapping.

Fig. 12. Mobile walking robot.

Fig. 11 depicts an advanced robotic platform viewed from a top-down perspective, showcasing a complex assembly of mechanical and electronic components designed for multifunctional tasks. The robot is built on a rigid, rectangular base that supports a variety of sensors and actuators. Central to the structure is a vertical post mounted with a digital display, presumably serving as an interface for monitoring and control. The arrangement of components indicates a modular design, with multiple actuation units at the corners equipped with what appears to be servo motors, suggesting capabilities for precise movement control and adjustment.

The visible electronics and wiring underscore the robot's sophistication, with circuit boards and a tangle of wires that imply a high degree of integration and connectivity among the various subsystems. The platform is likely intended for indoor applications, as suggested by the clean and flat surface on which it is stationed. This robot could be utilized for experiments that require stability and precision, such as those in automation, material handling, or advanced manufacturing settings. The inclusion of multiple sensors and actuators hints at the robot's potential for complex tasks involving manipulation, positioning, and environmental interaction, reflecting a high level of engineering investment aimed at versatility and performance in controlled environments.

Fig. 12 presents a sophisticated robotic platform designed for precision and adaptability across varied operational settings. Mounted on a robust aluminium frame, the robot features several articulated limbs and joints, hinting at its capability for performing complex tasks and maneuvers. At the core of the platform, a vertical column rises to support a digital display and control unit, tools likely essential for monitoring operations in real-time and making necessary adjustments. This system is further augmented by a variety of sensors and devices, including a LiDAR sensor, which suggests its utility in spatial mapping and autonomous navigation.

The robot is depicted in an indoor laboratory environment, underscored by a meticulously organized background and structured flooring that suggest its utilization in controlled experimental or developmental activities. The scene is replete with an array of visible electronic components and a complex network of wiring, highlighting the robot's sophisticated technological integration. This elaborate configuration is not merely for complexity's sake but serves a practical purpose by endowing the robot with a remarkable degree of adaptability and functionality. Such features make the robot immensely versatile, apt for a spectrum of applications extending from automated material handling to intricate assembly tasks in both research and industrial settings. This adaptability is critical in

modern industries where the demands and functionalities required can vary significantly. Consequently, the robot's design caters to a broad range of tasks, enabling it to perform with high efficiency and adapt quickly to new challenges, illustrating the cutting-edge of current robotic capabilities and the potential for future advancements in automation technology.

The robot's design strategically emphasizes modularity and flexibility, enabling rapid modifications and enhancements that are tailored to specific operational and research demands. This adaptability is critical, allowing the robotic platform to evolve alongside changing technological landscapes and varying project requirements, thus significantly enhancing its applicability in dynamic industrial and research environments. By facilitating such adaptability, the system not only extends the frontiers of robotic automation capabilities but also exemplifies the continuous innovation characterizing the field of robotic engineering and design. This design philosophy ensures that the robotic system remains a valuable asset in progressive applications, effectively responding to emerging challenges and opportunities in automation technology.

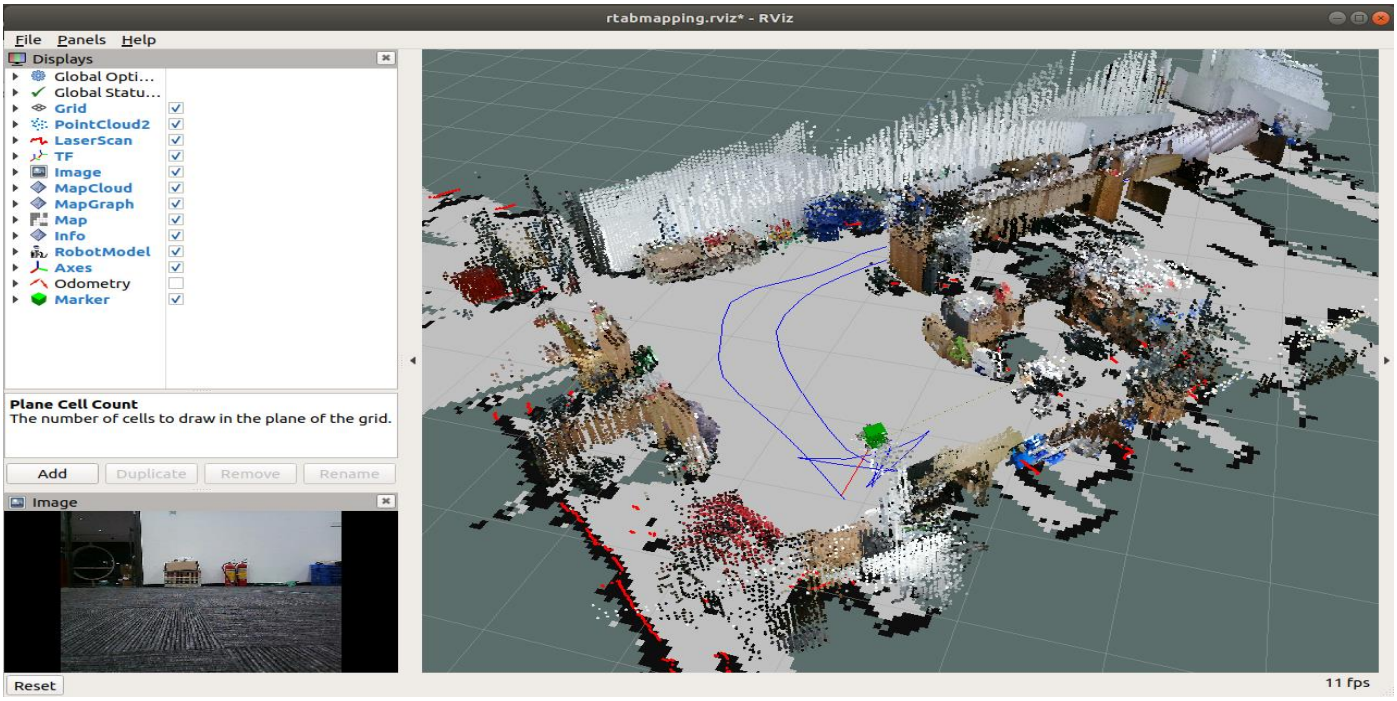


Fig. 13. 3D Simultaneous localization and mapping in mobile robots.

Fig. 13 illustrates the results of a 3D mapping process using advanced Simultaneous Localization and Mapping (SLAM) technology, displayed through the RViz visualization environment. The scene captures a comprehensive point cloud representation of an indoor environment, featuring a detailed and textured map constructed from data collected via sensors equipped on a mobile robot. The multicolored point clouds signify different objects and surfaces within the space, effectively distinguishing between floors, walls, furniture, and other objects based on their spatial and geometric properties.

The interface shown in the figure provides a clear view of various data layers and navigation paths. Blue lines depict the trajectory of the robot as it navigates through the environment, highlighting the path planning and movement executed during the mapping process. Each element in the RViz interface, such as point clouds, laser scans, and odometry data, is marked with checkboxes, allowing users to selectively view or hide different layers for better analysis and troubleshooting of the mapping data. The inset at the bottom left shows the real-world camera view corresponding to the robot's perspective, providing a ground-level context that complements the 3D spatial data. This integrated display aids in evaluating the accuracy and completeness of the SLAM process, essential for applications in autonomous navigation and robotic perception.

VI. Discussion

The development and implementation of sophisticated robotic platforms, as depicted in Fig. 12, illustrate significant advancements in the field of robotics and automation. These systems, equipped with articulated limbs, advanced sensor arrays, and modular frameworks, underscore a significant shift towards more adaptable and versatile robotic solutions. This section explores the implications of such technologies, focusing on their potential impact on industrial applications, challenges associated with their deployment, and the future trajectory of robotic systems design.

Adaptability and Application. The featured robotic platform is a prime example of the trend towards customization and flexibility in robotic system design. Equipped with a sturdy aluminum frame and multiple articulated limbs, the robot is capable of performing a wide array of tasks ranging from simple material handling to complex assembly operations [39]. The central integration of a digital display and control unit on a vertical column facilitates real-time monitoring and

adjustments, which are crucial in dynamic industrial environments. The inclusion of LiDAR sensors enhances the robot's navigation and spatial awareness capabilities, enabling it to perform tasks in unstructured environments that were traditionally challenging for automated systems [40]. Such versatility is increasingly demanded in industries where customization and adaptability to varied operational contexts are required.

Technological Integration and System Complexity The integration of diverse technologies, including advanced sensors and intricate wiring systems, brings about enhanced functionality but also introduces complexity in the maintenance and operation of these robots. The intricate array of components requires sophisticated diagnostic tools and skilled personnel for maintenance, which can increase operational costs [41]. Furthermore, the complexity may impact the system's robustness, as more components can lead to increased points of failure. However, the modular design approach mitigates some of these challenges by enabling easier upgrades and replacements, thus prolonging the system's operational life and adapting to evolving technological advances without necessitating complete system overhauls.

Challenges in Deployment. Deploying such advanced robotic systems in real-world industrial settings poses several challenges [42]. First, the initial cost of investment and integration into existing systems can be significant. Industries looking to adopt these technologies must consider not only the purchase and installation costs but also the training required for their workforce. Additionally, while the adaptability of the robot allows for its application in various settings, each new environment or task can require extensive reconfiguration and testing to ensure optimal performance. This adaptation process can consume time and resources, slowing down the integration process and potentially impacting production schedules.

Future Directions in Robotic Research and Development. Looking forward, the continuous evolution of robotic technologies promises even greater capabilities and more profound impacts on industrial and research applications. Future developments may focus on enhancing artificial intelligence and machine learning integrations, allowing robots to make more autonomous decisions based on real-time data analysis [43]. This advancement could lead to greater efficiency and precision in tasks such as predictive maintenance and complex decision-making processes. Additionally, the drive towards sustainable practices may influence robotic design, prioritizing energy efficiency and the use of environmentally friendly materials.

Moreover, the collaboration between robotics and other emerging technologies such as the Internet of Things (IoT) and augmented reality (AR) could further enhance the capabilities of robotic systems. For instance, IoT integration can enable a fleet of robots to communicate and operate in a coordinated manner, increasing productivity and efficiency. Meanwhile, AR can facilitate more intuitive interfaces for human-robot interaction, enhancing the usability of robotic systems in complex tasks.

VII. Conclusion

In conclusion, the exploration of advanced robotic platforms, as detailed in the provided analyses, underscores the significant strides made in the robotics and automation industry. The integration of complex sensor systems, articulated mechanical structures, and sophisticated control units within robust, modular frameworks exemplifies the technological evolution aimed at enhancing precision, adaptability, and functionality across various operational environments. These developments signify a pivotal shift towards more versatile automated solutions, capable of handling intricate tasks that require high levels of customization. The deployment of such technologies, despite the challenges associated with complexity and initial setup costs, promises substantial improvements in efficiency and operational capabilities for industries ranging from manufacturing to logistics. Looking forward, the continued advancement in robotics is expected to further merge with cutting-edge technologies like artificial intelligence, augmented reality, and the Internet of Things, broadening the scope of robotic applications and deepening their impact on industrial processes. This integration will not only refine the capabilities of robotic systems in terms of autonomy and decision-making but also enhance their interaction with human operators, thereby catalyzing a new era of innovation in automation that could fundamentally reshape industry standards and operational paradigms. As this field progresses, it will undoubtedly offer new opportunities and challenges, driving forward the capabilities and understanding of robotic systems in complex, dynamic environments.

Acknowledgment

This study was funded by the research project BR20280990 development, development of methods for solving fundamental problems of fluid and gas mechanics, new deformable bodies, and reliability of machines, mechanisms, robotics, energy efficiency.

References

[1] Panigrahi, P. K., & Bisoy, S. K. (2022). Localization strategies for autonomous mobile robots: A review. Journal of King Saud University-Computer and Information Sciences, 34(8), 6019-6039.

[2] Shaikh, B., Faraz, S. M., Jafri, S. R., & Ali, S. U. (2021). Self-localization of mobile robot using map matching algorithm. J. Eng. Appl. Sci, 40(1), 69-77.

[3] Zeki, I. M., Hashim, M. A., Hammood, M. M., & Nori, R. (2022). A Review on Localization Algorithms of Mobile Robot in Different Environments. Journal of Algebraic Statistics, 13(2), 3555-3580.

[4] Faisal, M., & ElGibreen, H. (2021, April). Adaptive Self-Localization System for Low-Cost Autonomous Robot. In 2021 7th International Conference on Control, Automation and Robotics (ICCAR) (pp. 114-121). IEEE.

[5] Ullah, I., Adhikari, D., Khan, H., Anwar, M. S., Ahmad, S., & Bai, X. (2024). Mobile robot localization: Current challenges and future prospective. Computer Science Review, 53, 100651.

[6] Kammel, C., Kögel, T., Gareis, M., & Vossiek, M. (2022). A cost-efficient hybrid UHF RFID and odometry-based mobile robot self-localization technique with centimeter precision. IEEE Journal of Radio Frequency Identification, 6, 467-480.

[7] Altayeva, A., Omarov, B., & Im Cho, Y. (2018, January). Towards smart city platform intelligence: PI decoupling math model for temperature and humidity control. In 2018 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 693-696). IEEE.

[8] Lin, H. Y., & He, C. H. (2021). Mobile robot self-localization using omnidirectional vision with feature matching from real and virtual spaces. Applied Sciences, 11(8), 3360.

[9] Altayeva, A., Omarov, B., & Im Cho, Y. (2017, December). Multi-objective optimization for smart building energy and comfort management as a case study of smart city platform. In 2017 IEEE 19th International Conference on High Performance Computing and Communications; IEEE 15th International Conference on Smart City; IEEE 3rd International Conference on Data Science and Systems (HPCC/SmartCity/DSS) (pp. 627-628). IEEE.

[10] Huang, J. (2024, June). Autonomous navigation and path planning for mobile robots based on SLAM. In AIP Conference Proceedings (Vol. 3144, No. 1). AIP Publishing.

[11] Ding, H., Zhang, B., Zhou, J., Yan, Y., Tian, G., & Gu, B. (2022). Recent developments and applications of simultaneous localization and mapping in agriculture. Journal of field robotics, 39(6), 956-983.

[12] Li, Z., Zhao, C., Wang, J., Hou, X., Hu, J., Pan, Q., & Jia, C. (2022). High-accuracy robust SLAM and real-time autonomous navigation of UAV in GNSS-denied environments. In Advances in Guidance, Navigation and Control: Proceedings of 2020 International Conference on Guidance, Navigation and Control, ICGNC 2020, Tianjin, China, October 23–25, 2020 (pp. 1099-1108). Springer Singapore.

[13] Hani, U., & Moin, L. (2021). Realtime autonomous navigation in V-Rep based static and dynamic environment using EKF-SLAM. Int J Rob & Autom ISSN, 2722(2586), 2586.

[14] Meng, J., Wang, S., Jiang, L., Hu, Z., & Xie, Y. (2023). Accurate and efficient self-localization of AGV relying on trusted area information in dynamic industrial scene. IEEE Transactions on Vehicular Technology, 72(6), 7148-7159.

[15] Lee, Y. C. (2024). LSMCL: Long-term Static Mapping and Cloning Localization for autonomous robot navigation using 3D LiDAR in dynamic environments. Expert Systems with Applications, 241, 122688.

[16] Rico, F. M., Hernández, J. M. G., Pérez - Rodríguez, R., Peña - Narvaez, J. D., & Gómez - Jacinto, A. G. (2024). Open source robot localization for nonplanar environments. Journal of Field Robotics, 41(6), 1922-1939.

[17] Polvara, R., Del Duchetto, F., Neumann, G., & Hanheide, M. (2021). Navigate-and-seek: a robotics framework for people localization in agricultural environments. IEEE Robotics and Automation Letters, 6(4), 6577-6584.

[18] Wang, X., Mizukami, Y., Tada, M., & Matsuno, F. (2021). Navigation of a mobile robot in a dynamic environment using a point cloud map. Artificial Life and Robotics, 26, 10-20.

[19] Sadeghi Esfahlani, S., Sanaei, A., Ghorabian, M., & Shirvani, H. (2022). The deep convolutional neural network role in the autonomous navigation of mobile robots (SROBO). Remote Sensing, 14(14), 3324.

[20] Balaska, V., Bampis, L., & Gasteratos, A. (2022). Self-localization based on terrestrial and satellite semantics. Engineering Applications of Artificial Intelligence, 111, 104824.

[21] Lu, Y., Ma, H., Smart, E., & Yu, H. (2021). Real-time performance-focused localization techniques for autonomous vehicle: A review. IEEE Transactions on Intelligent Transportation Systems, 23(7), 6082-6100.

[22] Chang, Y., Cheng, Y., Manzoor, U., & Murray, J. (2023). A review of UAV autonomous navigation in GPS-denied environments. Robotics and Autonomous Systems, 104533.

[23] Tang, Y., Zhao, C., Wang, J., Zhang, C., Sun, Q., Zheng, W. X., ... & Kurths, J. (2022). Perception and navigation in autonomous systems in the era of learning: A survey. IEEE Transactions on Neural Networks and Learning Systems, 34(12), 9604-9624.

[24] Alkendi, Y., Seneviratne, L., & Zweiri, Y. (2021). State of the art in vision-based localization techniques for autonomous navigation systems. IEEE Access, 9, 76847-76874.

[25] Sasaki, M., Tsuda, Y., & Matsushita, K. (2024). Development of Autonomous Mobile Robot with 3DLidar Self-Localization Function Using Layout Map. Electronics 2024, 13, x. https. doi. org/10.3390/xxxxx Academic Editors: Quan Qian and Xing Wu Received, 26.

[26] Liu, Y., Wang, S., Xie, Y., Xiong, T., & Wu, M. (2024). A Review of Sensing Technologies for Indoor Autonomous Mobile Robots. Sensors, 24(4), 1222.

[27] McGuire, J. L., Law, Y. W., Doğançay, K., Ho, S. Y., & Chahl, J. (2022). Optimal maneuvering for autonomous vehicle self-localization. Entropy, 24(8), 1169.

[28] Martín Rico, F., Guerrero Hernández, J. M., Pérez-Rodríguez, R., Peña-Narvaez, J. D., & García Gómez-Jacinto, A. (2024). Open source robot localization for nonplanar environments.

[29] Charroud, A., El Moutaouakil, K., Palade, V., Yahyaouy, A., Onyekpe, U., & Eyo, E. U. (2024). Localization and Mapping for Self-Driving Vehicles: A Survey. Machines, 12(2), 118.

[30] Fusic, S., & Sugumari, T. (2023). A review of perception-based navigation system for autonomous mobile robots. Recent Patents on Engineering, 17(6), 13-22.

[31] Niloy, M. A., Shama, A., Chakrabortty, R. K., Ryan, M. J., Badal, F. R., Tasneem, Z., ... & Saha, D. K. (2021). Critical design and control issues of indoor autonomous mobile robots: A review. IEEE Access, 9, 35338-35370.

[32] Joshi, R., Bhaiya, D., Purkayastha, A., Patil, S., & Deshpande, A. (2021, August). Simultaneous Navigator for Autonomous Identification and Localization Robot. In 2021 IEEE Region 10 Symposium (TENSYMP) (pp. 1-6). IEEE.

[33] Jeya, R., Vanidhasri, R., Manikanthan, S. V., Kishore, K. H., & Manikandan, S. (2024). Indoor Environment Surveillance via an Autonomous Mobile Robotic Device. Nanotechnology Perceptions, 584-610.

[34] Wu, Z., Yue, Y., Wen, M., Zhang, J., Yi, J., & Wang, D. (2021). Infrastructure-free hierarchical mobile robot global localization in repetitive environments. IEEE Transactions on Instrumentation and Measurement, 70, 1-12.

[35] Feng, S., Shi, H., Huang, L., Shen, S., Yu, S., Peng, H., & Wu, C. (2021). Unknown hostile environment-oriented autonomous WSN deployment using a mobile robot. Journal of Network and Computer Applications, 182, 103053.

[36] Farag, W. (2022). Self-driving vehicle localization using probabilistic maps and Unscented-Kalman filters. International Journal of Intelligent Transportation Systems Research, 20(3), 623-638.

[37] Alam, M. S., Mohamed, F. B., & Hossain, A. B. (2024). Self-Localization of Guide Robots Through Image Classification. Baghdad Science Journal, 21(2 (SI)), 0832-0832.

[38] Guo, Y., Xie, Y., Chen, Y., Ban, X., Sadoun, B., & Obaidat, M. S. (2022). An efficient object navigation strategy for mobile robots based on semantic information. Electronics, 11(7), 1136.

[39] Buratowski, T., Garus, J., Giergiel, M., & Kudriashov, A. (2022). Real-time 3d mapping in isolated industrial terrain with use of mobile robotic vehicle. Electronics, 11(13), 2086.

[40] Jiao, H., & Chen, G. (2023). Global self-localization of redundant robots based on visual tracking. International Journal of System Assurance Engineering and Management, 14(2), 529-537.

[41] Omarov, B., Altayeva, A., & Cho, Y. I. (2017). Smart building climate control considering indoor and outdoor parameters. In Computer Information Systems and Industrial Management: 16th IFIP TC8 International Conference, CISIM 2017, Bialystok, Poland, June 16-18, 2017, Proceedings 16 (pp. 412-422). Springer International Publishing.

[42] Kameyama, S., Okumura, K., Tamura, Y., & Défago, X. (2021, May). Active modular environment for robot navigation. In 2021 IEEE International Conference on Robotics and Automation (ICRA) (pp. 8636-8642). IEEE.

[43] Wang, Y., Fan, Y., Wang, J., & Chen, W. (2024). Long-term navigation for autonomous robots based on spatio-temporal map prediction. Robotics and Autonomous Systems, 179, 104724.

Optimizing LSTM-Based Model with Ant-Lion Algorithm for Improving Thyroid Prognosis

Maria Yousef
Department of Computer Science, University of Petra, Amman, Jordan

***Abstract*—In the healthcare sector, early and accurate disease detection is essential for providing appropriate care on time. This is especially crucial in thyroid problems, which can be difficult to diagnose because of their many symptoms. This study aims to propose a new thyroid disease prediction model by utilizing the Ant Lion Optimization (ALO) approach to enhance the hyperparameters of the Long Short-Term Memory (LSTM) deep learning algorithm. To achieve this, after the preprocessing step, we utilize the entropy technique for feature selection, which selects the most important features as an optimal subset of features. The ALO is then employed to optimize the LSTM, identifying the optimal hyperparameters that can influence the model and enhance its efficiency. To assess the suggested methodology, we chose the widely used thyroid disease data. This dataset contains 9,172 samples and 31 features. A set of criteria was used to evaluate the model's performance, including accuracy, precision, recall, and F1 score. The experimental results showed that: 1) the entropy technique in the feature selection step can reduce the total number of features from 31 to 10; 2) the recommended strategy, which selected the optimal hyperparameter for the LSTM using the Alo algorithm, improved the classifier overall by 7.2% and produced the highest accuracy of 98.6%.**

Keywords*—Thyroid disease; LSTM; ALO; prediction model; optimization algorithm

I. Introduction

Incidences of thyroid illness have increased recently. The thyroid gland has one of the most significant functions in controlling metabolism. The thyroid gland secretes levothyroxine (T4) and triiodothyronine (T3), and irregularities in the gland can result in a variety of disorders [1]. Inadequate thyroid hormones can cause hyperthyroidism and hypothyroidism, the most prevalent disorders. If individuals with thyroid disease do not manage their metabolism and body weight appropriately, they may experience severe problems such as reduced taste, diminished smell, sadness, and memory loss. According to the World Health Organization, 20 million people have had some form of thyroid disease in the past 2 years [2]. Moreover, thyroid disorders affect 12% of the population at least once in their lives. Statistics indicate that we should not disregard thyroid-related illnesses lightly. As a result, the use of prediction systems for thyroid disease detection has grown in significance and is currently a field of research.

Health care professionals are always concerned with finding a cure for illnesses, so making an accurate diagnosis when a patient needs it is crucial. In the traditional way, only an experienced doctor is able to thoroughly assess the situation when predicting a thyroid problem; otherwise, it can be a challenging procedure that results in an incorrect diagnosis. A proactive prediction of thyroid illness is critical to appropriately treating the patient at the appropriate time, saving lives and medical [3].

Current data processing and computer technologies have made it possible to identify different forms of thyroid illness, such as hyperthyroidism and hypothyroidism, and forecast thyroid disease early on. These advances have also made machine learning and deep learning approaches more accessible. Most existing research efforts focus on binary classification issues, categorizing individuals as either thyroid sufferers or healthy. In contrast, there are very few multiclass-based detection methods. Prior research in this area also concentrated on directly implementing machine learning algorithms and contrasting their outcomes, regardless of the attributes employed, their significance, and their quantity, which impacts the models' accuracy and gives rise to the high-dimensionality problem. Despite the high accuracy of reporting techniques, their evaluation on less than 1000 samples yields inconsistent results. To address these concerns, this study presents the following contributions:

- Proposed a new thyroid prediction model that employs a long-short-term memory (LSTM) deep learning algorithm, bolstered by the Ant Lion Optimization (ALO) metaheuristic algorithm. This model serves as an enhanced machine learning-based method for predicting thyroid illness, specifically targeting a multi-class problem.
- Handling the high dimensionality problem and selecting the optimal subset of features that can enhance the model's performance through machine learning-based feature selection using an additional tree classifier.
- Demonstrate the impact of the ALO optimization algorithm on the performance of the LSTM classifier through the suggested experiments.

The rest of the article is structured as follows: Section II summarises the previous attempts to identify and classify thyroid-related illnesses. Section III provides a background on all of the algorithms and techniques used in this study. Section IV presents the suggested approach to solving the thyroid illness prediction issue. Section V describes the experimental results we achieved in our investigation, while Section VI presents the study's conclusion along with our inputs.

II. Related Work

This section covers summaries of many machines learning based thyroid disease prediction systems.

A study by [4] suggested a way to predict thyroid disease that uses five different classification algorithms: Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Multilayer Perceptron, and Naive Bayes. The authors of this paper gathered the dataset from the Iraqi population, which included 1250 male and female individuals ranging in age from 1 year to 90 years. It includes three types of classes: hypothyroidism, normal, and hyperthyroidism—as well as 17 attributes that represent patient information and relevant medical analyses. A comparison and implementation of several algorithms performances was conducted. The result shows that RF outperformed other algorithms in the ability to predict; it achieved 98% accuracy.

In an effort to improve the effectiveness of the KNN algorithm in thyroid prediction, Abbad Ur Rehman et al. [5] examined two well-known feature selection methods: chi-square feature selection and L1-norm feature selection. This study utilized two datasets for training one from KEEL dataset repository that contains 7200 cases, with 13 features and another from a registered hospital in Pakistan was used for testing. It contains 3300 instances and 11 features. After assessing the performance of unique distance functions of KNN, Euclidean and Cosine distance functions obtained 97% accuracy at the k value of 1 with the use of chi-square-based feature selection approach for new suggested dataset.

A comparative study on thyroid illness detection was conducted by [6]. In this study, the authors utilizing a multilayer neural network, a learning vector quantization-based neural network, and a probability-based neural network. This study uses the UCI machine learning database's thyroid illness dataset. There are five features and 215 samples in total. The experiment's findings demonstrated that neural network structures can effectively assist in thyroid illness diagnosis; the probability-based neural network outperformed other neural networks, achieving an accuracy rate of 94.8%.

Ahmad et al. in [7] described a mixed decision support system that uses k nearest-neighbour (kNN) weighted preprocessing, an adaptive neurofuzzy inference system (ANFIS), and linear discriminant analysis (LDA). The LDA-kNN-ANFIS approach starts with LDA, which lowers the illness dataset dimensionality and removes pointless features. The second step uses a kNN-based weighted preprocessor to preprocess a subset of the attributes. In the final step, the adaptive neurofuzzy inference system uses preprocesses, chosen characteristics as an input for diagnosis. The suggested technique tested the system's overall performance using a thyroid illness dataset from the University of California, Irvine's machine learning repository. The calculated classification analysis using this approach's accuracy, sensitivity, and specificity values came out to be 98.5, 94.7, and 99.7%, in that order.

III. BACKGROUND

A. Ant Lion Optimization (ALO)

It is a metaheuristic optimizer that draws inspiration from nature, the ALO algorithm emulates the natural predatory behaviour of ant lions [8]. The algorithm aims to resolve complex optimization issues by emulating the communication between antlions, representing superior solutions, and ants, representing potential answers. By building traps to catch other ants, antlions direct the hunt for the best answers [9]. The ALO algorithm successfully explores the search space by using a random walk mechanism for ants that is impacted by the placement of antlions. The ALO algorithm guarantees fast convergence to high-quality solutions and thorough coverage of the solution space by striking a balance between exploration and exploitation. To direct the search for the best answers, it makes use of heuristic data and pheromone trails [10], where the probability of the pheromone update is defined as follows:

$$P_i = \left[\frac{F_i^n}{\sum_i F_i^n}\right] \quad (1)$$

Where:

P_i: is a possibility that any path will be selected.

F_i^n: is the strength of the pheromone on every path.

The ALO algorithm is divided into two populations: an ant population and an antlion population. The general steps that ALO employs to alter these two sets and get the global optimum for a specific optimization issue are as follows:

- The primary search agents of the ALO are the Ant set, which is first initialized with random values.
- In every iteration, the fitness value of every ant is assessed using an objective function.
- Ants use random walks to get around the antlions in the search area.
- There is never an assessment of the antlion population. In actuality, antlions moved to the new placements of ants in subsequent iterations if the ants improved from their initial position, which they presumed to be on.
- Each ant has a single antlion, which updates its location whenever the ant becomes fitter.
- Additionally, there is an elite antlion that affects ant movement regardless of distance.
- Any antlion that surpasses the elite will be replaced by the elite.
- Once an end condition is satisfied, steps b through g are repeated.
- The elite antlion's location and fitness value are given back as the most accurate estimate of the global optimum.

By repeating this procedure, the ALO algorithm efficiently converges to optimum or nearly optimal solutions for a range of optimization issues. Fig. 1 displays the ALO algorithm flowchart:

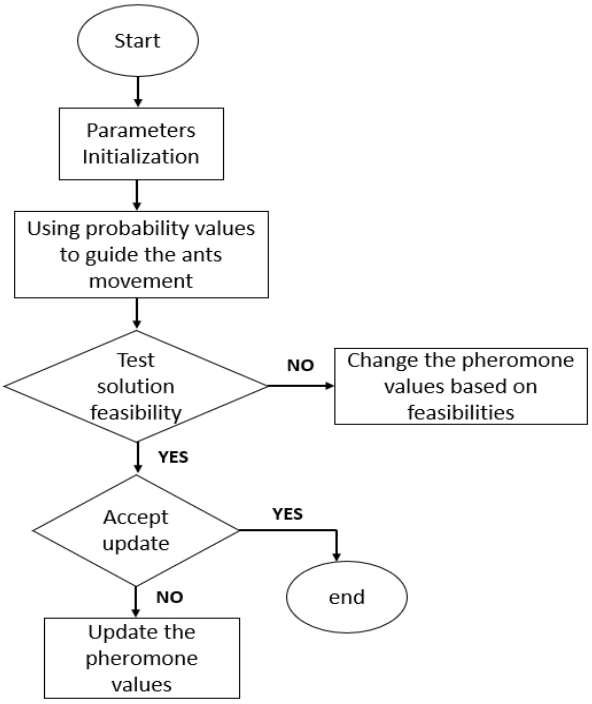


Fig. 1. ALO flowchart.

B. Long Short Term Memory (LSTM)

A unique kind of recurrent neural network (RNN) architecture was created to handle sequential input and overcome the shortcomings of conventional RNNs in identifying long-term relationships [11]. Long-term state memory (Long-term Memory) is achieved by LSTM networks through the use of memory cells that hold onto information over extended periods of time [12]. Each cell in LSTM consists of three gates: the input gate, the forget gate, and the output gate. These gates control the information flow by selecting which input elements to keep, which to reject, and which to go on to the subsequent time step [13]. For applications requiring time series, prediction processing, and other sequential data, LSTM networks are especially useful because they can learn to remember pertinent information and discard irrelevant data by dynamically modifying these gates.

The operation of an LSTM can be explained through the repetitive process that follows by each element in an ordered series. For each time step, the input gate determines how much a new input should impact the cell's current state. Conversely, the forget gate decides whether to remember parts of the previous cell state [14]. The output gate then generates the current output by merging the updated cell state with the new input. This output is not only forwarded to the following layer or ultimate prediction but also returned to the cell in order to analyze the subsequent element in the sequence. This approach enables LSTMs to learn relationships spanning several time steps, which is essential for modelling sequential data well. It also lets LSTMs keep an updated version of short-term memory at each step [15].

In traditional RNN, this repeating unit has a straightforward structure, such as the Tanh layer. While the LSTM and repeating modules have the same design, they differ in their architecture [16]. Rather than having a single neural network layer, Fig. 2 shows four layers that interact in a very special way.

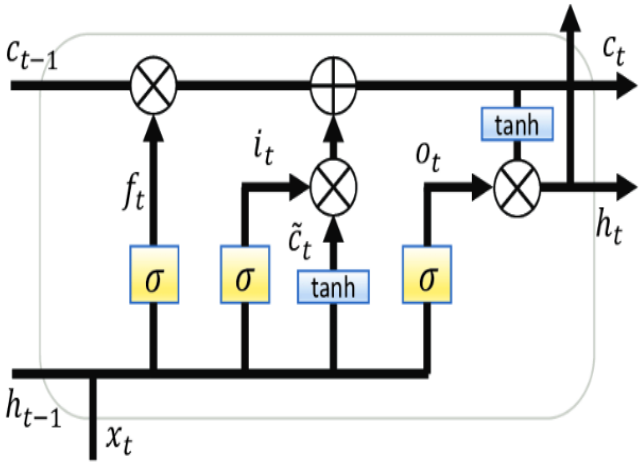


Fig. 2. LSTM architecture.

Where, h_t stands for the current layer's output, x_t for the current input, and h_{t-1} for the output of the previous layer. C_t indicates the current layer's cell status. While the symbols C_{t-1}, ft, and $\tilde{C}_t$ represent the cell state, forget gate layer, and candidate vector, respectively.

We can establish the system by using the forget gate layer given by Eq. (2):

$$f_t = \sigma(W_f . [h_{t-1}, x_t] + b_f) \quad (2)$$

Where, σ: is the sigmoid function, W_f :denotes the forget layer's weights, and b_f: denotes the forget layer's bias. Next, we employ the following Eq. (3), and (4) to update the input layer:

$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\tilde{C}_t = tanh(W_c . [h_{t-1}, x_t] + b_c) \quad (4)$$

Where, i_t : denotes the output of the input gate layer, W_i: denotes the weights of the input gate layer, b_i : indicates its bias, the output of the input candidate layer is represented by $\tilde{C}_t$. Tanh is the output obtained from the input candidate layer, W_c stands for the hyperbolic tangent function, and b_c is the candidate layer's bias. Next, we employ the following Eq. (5) to defined the output layer:

$$o_t = \sigma(W_o . [h_{t-1}, x_t] + b_o) \quad (5)$$

Where, the output of the output gate layer is represented by the symbol o_t. W_o, and $\boldsymbol{b_o}$, which stand for output gate weights and bias, respectively. Eq. (6) provides the output of the output gate layer. Eq. (6) yields the output of the output gate layer, and Eq. (7) provides the current cell state.

$$h_t = o_t * tanh(C_t) \quad (6)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t) \quad (7)$$

IV. Research Methodology

Fig. 3 illustrates the five primary steps of the suggested early diagnostic model. The first stage explains the Thyroid disease dataset. Preprocessing is the second stage, where the main goals are analysis, resizing, and enhancing the quality of the dataset. Using the decision tree entropy technique, features are selected in the third step to create the best possible subset of data. The LSTM-based model uses these as training vectors and then employs a meta-heuristic technique to maximize convergence and improve performance. Ultimately, a range of assessments will be conducted to classify the illness classes and evaluate the suggested system utilizing a variety of criteria.

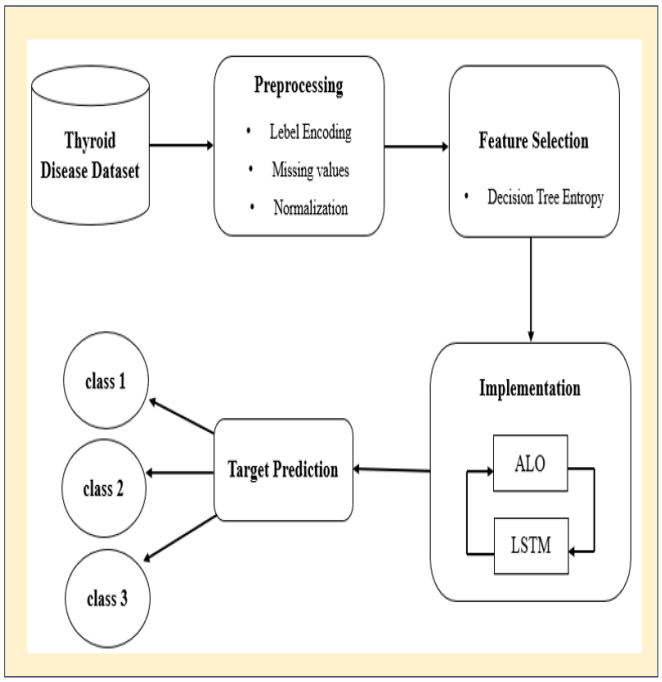


Fig. 3. The proposed system schema.

A. Thyroid Disease Dataset

This investigation used the thyroid dataset from the University of California, Irvine (UCI) repository [17]. We chose this particular data set because it is well-established in the field of thyroid illness classification from earlier studies. The 9,172 samples and 31 features in the dataset record a variety of medical characteristics and thyroid-related test findings. The main features contain patient demographic traits like sex and age, and multiple medical necessities and treatments such as sick, on_thyroxine, thyroid_surgery, TSH measurement, and pregnancy. The dataset also includes various thyroid hormone levels (TSH, T3, TT4, T4U, FTI, and TBG), with some missing values. Additionally, it records the referral source and a target variable, indicating the outcome or diagnosis. Additionally, the dataset contains many thyroid hormone levels with some missing data (TT4, T3, TSH, TT4, FTI, TBG, and T4U). It also logs the referral source and a target variable that indicates the diagnosis or outcome. This dataset is useful for analyzing thyroid diseases because it provides a comprehensive view of patient health and thyroid-related medical history. Table I displays a description of each feature.

TABLE I. ASD Dataset Features

Features	Description	Type
age	The patient age	integer
sex	Identifies a sex of patient	String
on_thyroxine	Whether the patient is taking thyroxine medication	Bool
on_antithyroid_meds	If the patient takes antithyroid medication	
query on thyroxine	Did the patient performed a thyroxine test?	
sick	Does the patient suffer from other diseases?	
pregnant	Whether patient is pregnant	
thyroid_surgery	Whether patient has undergone thyroid surgery	
I131_treatment	Whether the patient is receiving therapy with I131	
query_hypothyroid	Whether the person thinks they are hypothyroid	
lithium	If the patient takes lithium medication	
goitre	Whether the patient suffers from goitre	
tumor	If the patient has a tumour	
hypopituitary	Whether the patient has an overactive pituitary gland	Float
psych	If the patient takes psychiatric or psychological medication	Bool
TSH_measured	Whether TSH was measured in the blood	
TSH	TSH level in blood from lab work	Float
T3_measured	Whether a blood test for T3 was performed	Bool
T3	T3 level in blood from lab work	Float
TT4_measured	Whether TT4 was measured in the blood	Bool
TT4	T4 level in blood from lab work	Float
T4U_measured	Whether a blood test for T4U was performed	Bool
T4U	T4U level in blood from lab work	Float
FTI_measured	Whether a blood test for FTI was performed	Bool
FTI	FTI level in blood from lab work	Float
TBG_measured	If blood TBG levels were tested	Bool
TBG	TBG level in blood from lab work	float
referral_source	Refers to the healthcare facility or physician from whom the patient was referred for additional assessment, diagnosis, or treatment.	String
target	Final diagnosis	String
patient_id	Unique patient ID	

B. Preprocessing

The information prehandling process is regarded as a critical step in data mining tasks. Crude information frequently contains missing or erroneous data. This leads to a state of confusion in machine learning forecasting. This step includes cleaning, extracting, and modifying data into a suitable format for machine processing [18].

1) Label encoding: The categorical variables in the thyroid illness dataset used in this study consist of categorical values that must be converted into a quantitative form. The data is transformed into certain numerical forms using the one-hot encoding technique [19]. This labelling procedure assigns a value of 0 or 1 to each data point based on its category, enabling machine learning algorithms to differentiate between the two groups. During this phase, we convert data for features such as "sex", "sick", etc. from Boolean type values (yes/no) or (t/f) to binary values (1/0), as shown in Fig. 4.

1: age Numeric	2: sex	3: on_thyroxine	4: query_on_thyroxine	5: on_antithyroid_meds	6: sick	7: pregnant	8: thyroid_surgery	9: I131_treatment	10: query_hypothyroid
29.0	0	1	0	0	0	0	0	1	
29.0	0	1	0	0	0	0	0	0	
41.0	1	1	0	0	0	0	0	0	
36.0	1	1	0	0	0	0	0	0	
32.0	1	1	1	0	0	1	0	0	
60.0	1	0	1	0	0	0	1	0	
77.0	1	0	1	0	0	0	1	0	
28.0	0	0	1	0	0	0	1	0	
28.0	0	1	0	0	0	1	1	0	
28.0	0	0	1	0	0	1	0	0	
54.0	1	0	0	0	0	1	0	0	
42.0	1	0	1	0	0	1	0	0	
51.0	1	1	0	0	0	1	0	0	
51.0	1	0	0	0	0	0	0	0	
37.0	0	1	0	0	0	0	0	0	
16.0	1	0	0	0	0	1	0	0	

Fig. 4. Part of dataset after levelling.

2) Missing values: Missing data is a common issue, especially in the health industry. Almost every feature and patient record includes some missing data. Thyroid disease dataset has missing values for a number of features, including "TSH," "goitre," "TT4", and "T3." 3. Removing samples or features that have missing values is ineffective because it may lead to the loss of important information during the diagnosis process [20]. At this stage, we apply a filter approach that uses the mean of the training data to fill in the gaps and eliminate missing values from the dataset.

3) Normalization: The process of converting data into a suitable format that may be used for mining is known as data transformation. One method of data scaling is normalization, which is the act of bringing attribute values into a specified range [21]. The normalization strategy has been applied prior to the feature selection and modelling stages since varying attribute values make it more difficult to compare attributes and reduce the learning capacity of algorithms. Therefore, this study applied the Min-Max normalization approach to numeric data types. Each feature's greatest value is changed to 1, the lowest value to 0 [22], and all other values are transformed to a decimal interval. Fig. 5 shown the part of dataset after normalization.

22: TT4 Numeric	23: T4U_measured String	24: T4U Numeric	25: FTI_measured String	26: FTI Numeric	27: TBG_measured String	28: TBG Numeric
0.2006688963210...	0	0.47685185185185...	0	0.07457935425...	0	0.0645322661...
0.2107023411371...	0	0.63888888888888...	0	0.05979990904...	0	0.0595297648...
0.1989966555183...	0	0.7546296296296...	0	0.06093678944...	1	0.0545272636...
0.1989966555183...	0	0.24537037037037...	0	0.04956798544...	1	0.1295647823...
0.0903010033444...	0	0.37499999999999...	0	0.08367439745...	1	0.1795897948...
0.1086956521739...	0	0.18055555555555...	0	0.05297862664...	1	0.1295647823...
0.0535117056856...	0	0.28240740740740...	0	0.07457935425...	1	0.2696348174...
0.1906354515050...	0	0.66203703703703...	0	0.08140063665...	0	0.1145572786...
0.1237458193979...	0	0.64351851851851...	0	0.08367439745...	0	0.2146073036...
0.1354515050167...	0	0.28240740740740...	0	0.08481127785...	0	0.1545772886...
0.2190635451505...	0	0.14351851851851...	0	0.08253751705...	0	0.1345672836...
0.1722408026755...	0	0.27777777777777...	0	0.09163256025...	0	0.1445722861...
0.1638795986622...	0	0.28240740740740...	0	0.07457935425...	0	0.1295647823...
0.1906354515050...	0	0.09259259259259...	0	0.10982264665...	0	0.1595797898...
0.2023411371237...	0	0.66203703703703...	0	0.04956798544...	0	0.1545772886...

Fig. 5. Part of dataset after normalization.

C. Feature Selection

The information prehandling process is regarded as a critical step in data mining tasks. Crude information frequently contains missing or erroneous data. This leads to a state of confusion in machine learning forecasting. This step includes cleaning, extracting, and modifying data into a suitable format for machine processing [18].

The extra tree classifier [25], using the training dataset, generates numerous decision trees at random. After dividing the decision tree's nodes, we apply either the entropy criteria or the Gini index. In our feature selection procedure, we used entropy measures [26]. The entropy is calculated using Eq. (8). The proportion of rows in the dataset with the label i is represented by pi, while the number c denotes the unique class labels.

$$Entropy(E) = \sum_{i=1}^{c} p_i log_2(p_i) \quad (8)$$

Entropy evaluates the information about the condition of the features with respect to the target. We calculate the entropy of the features derived from each choice, and identify the significant features by computing the cumulative entropy values for each feature. Referred to as the shortlisted features, this group of traits has a high entropy. Fig. 6 uses Multilevel Feature Selection (MLFS) [27] to illustrate the significance of the features. With n_estimators = 300 and max_depth = 21, we employed an ETC classifier for MLFS, which rated the features according to their value. In this step, entropy was able to reduce the number of features from 30 to 10. The subset of features with scores greater than 0.02 is selected to train the learning models, namely 'Thyroid_surgey', 'pregnant', 'tumor', 'psych','TSH','T3', 'TT4', 'T4U', 'FTI' and 'TBG'.

D. Implementation

This stage involves using the Ant Lion Optimizer (ALO) as an optimizer for an LSTM model. During the optimization process, we may modify the algorithm movement equations to update the LSTM model weights and biases. The goal is to find the ideal values for the model's parameters to minimize the loss function and enhance performance.

There are numerous essential steps in the suggested approach when utilizing the ALO algorithm to optimize the hyperparameters of LSTM networks. The first step is establishing the hyperparameter search space for the LSTM model, which includes variables like learning rate, batch size, number of LSTM layers, and number of neurons. An "ant" represents every potential solution in this area, while "antlions"

symbolize the ideal or nearly ideal solutions. ALO creates an initial population of ants, each representing a collection of hyperparameters. A roulette wheel selection procedure guides ants towards antlions at the start of each iteration to replicate their exploration and exploitation of the solution space. The LSTM model assesses each ant's fitness based on its performance. Ants update their locations in the search space based on their interactions with antlions, and either keep or modify the top-performing solutions—antlions—as needed. This iterative procedure continues until convergence to ensure that the chosen hyperparameters effectively maximize LSTM performance. The following steps explain how to achieve this:

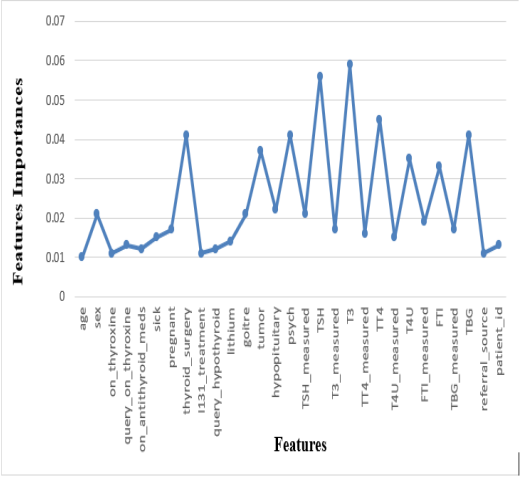


Fig. 6. Entropy value of the thyroid dataset features.

1) Initialization: Set random biases and weights at the beginning of the LSTM model. Determine how many antlions (M) and ants (N) are present. In the search area, place the ants at random starting points [28]. Set the antlions' starting positions at random throughout the search area. Decide on the maximum number of iterations (MaxIter).

2) Fitness evaluation: Use the objective function, which is usually the LSTM model's loss function, to determine each ant's and antlion's level of fitness [29].

3) Movement of ants: Every ant travels at random around the closest antlion inside a hypersphere. Ant movement is determined by the following formulas:

$$c_i(t) = Antlion_j(t) + l_c(t) \quad (9)$$

$$q_i(t) = Antlion_j(t) + h_q(t) \quad (10)$$

Where:

$c_i(t)$ and $q_i(t)$: represent the ith ant's lowest and maximum values at time t $Antlion_j(t)$: represents the location of the jth antlion at t. $l_c(t)$: represents the variable with the lowest value at time t among all variables is represented by $h_q(t)$: represents the variable that has the highest value at time t among all variables.

4) Movement of antlions: Within its capture radius, each antlion moves towards the most fit ant. The following equation determines how antlions move:

$$Antlion_j(t+1) = Antlion_j(t) + rand(). (Antlion_j(t) - X_f) \quad (11)$$

where, $rand()$ is a random value between 0 and 1, and $-X_f$ is the location of the fittest ant inside the capture radius of the jth antlion

5) Building trap: Utilizing a roulette wheel selection process, the fittest antlions are chosen according to their fitness scores [30]. To update the LSTM model's weights and biases inside their capture radius, the chosen antlions are employed.

6) Update: The movement equations are used to update the positions of every ant and antlion [31]. The revised antlion locations are used to update the weights and biases of the LSTM model. The modified LSTM model is used to reassess the fitness of each ant and antlion.

7) Termination: The algorithm ends when it reaches the maximum number of iterations or satisfies a stopping condition. where N represents the number of ants, M stands for the number of antlions, MaxIter signifies the maximum number of iterations, and lc(t) represents the lowest value of all the variables at time t [32].

Algorithm 1, written in Python, describes a classifying model for this study. The configuration of the LSTM model includes 50 units and a 972 softmax activation function to facilitate multi-class classification. The code utilizes the ALO algorithm to optimize the LSTM model. Through the model's training step, the ALO algorithm uses specific hyperparameters, such as 150 iterations and 15 ant lions.

Algorithm 1: Pseudocode of LSTM

```
# Define the LSTM-based model
model = Sequential ()
model.add (LSTM (units=50, input_shape= (time_steps,
num_features)))
model.add (Dense (units=num_classes,
activation='softmax'))
# Define the Ant Lion Optimizer
optimizer=AntLionOptimizer (num_iter=150,
num_antlion=15,
num_dimensions=model. count_params ())
# Compile the model
model. compile (optimizer=optimizer,
loss='categorical_crossentropy', metrics=['accuracy'])
# Train the model
model.fit (x_train, y_train, epochs=100, batch_size=32)
# Make predictions
predictions = model. predict(x_test)
```

V. Result and Discussion

To perform our experiments, we used the following software and hardware requirements: an Intel Core i7-6500U CPU with 16 GB of RAM and a clock speed of 4 GHz.

In the testing stage, we assessed how successfully the ALO_ LSTM-based model classified various kinds of thyroid illness.

After training the model on 80% of the dataset, we tested 30% to evaluate its performance. A confusion matrix [33], which included metrics like true positive, true negative, false positive, and false negative, was used to determine the results. Metrics including accuracy, recall, precision, and F1-Score were employed to assess the performance of the classification model. The following four parameters can be used to calculate the metrics and measurements that are employed.

1) True positive (TP): denotes correctly predicted positive values; in other words, it indicates that both the predicted and actual class values are true.

2) True negative (TN): these are the accurately predicted negative values, which indicate that neither the expected nor the actual class value is false.

3) False positive (FP): in situations where the predicted class is true but the actual class is false.

4) False negative (PN): This occurs when the predicted class is false, but the actual class is true.

The accuracy, recall, precision, and F1 score may be determined based on the previously mentioned parameters. These can be expressed as stated using the following equations:

$$Accuracy = \frac{TP+TN}{TP+FP+TP+TN} \quad (12)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (13)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (14)$$

$$F1 - score = \frac{TP}{TP+FN} \quad (15)$$

A. First Experiment

This study conducted two experiments to examine how the ALO metaheuristic algorithm may enhance the LSTM-based model's performance. In the first experiment, we employed the LSTM algorithms directly, regardless of the AlO, and used 'Adam' as an optimizer parameter. In contrast, in the second experiment, we used the ALO optimizer to improve the accuracy of the classification model. Table II shows the parameters used for LSTM in the first experiment. Where he Fig. 7 present the confusion matrix values used in determined the model performance.

TABLE II. THE LSTM PARAMETER FOR FIRST EXPERIMENTAL OUTCOMES

Parameters	Values
Model	Sequential
Hidden layer	4
Hidden unites	50
Neurons per hidden layer	30,30,30,1
Type of layer	LSTM
Activation hidden layer	Softmax
Epoch	100
Optimizer	ALO
Time step	1

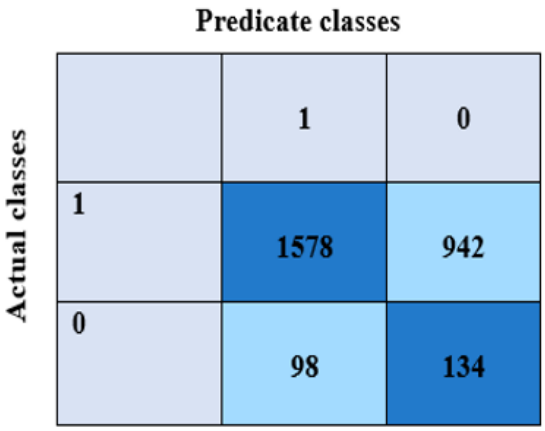


Fig. 7. Confusion matric of LSTM with Adam optimizer.

B. Second Experiment

The second experiment uses the ALO method as an optimizer for LSTM model parameters. Its purpose is to identify the optimum combination of weights and biases for the LSTM model, reducing the loss function while simultaneously speeding convergence through effective search space travel. The ALO algorithm employs specific hyperparameters, such as 150 iterations and 15 ant lions, during the model's training phase. The LSTM parameters for the second experiment are shown in Table III, and Fig. 8 details the confusion matrix from the second experiment. Additionally, Table IV displays the outcomes of our studies.

TABLE III. THE LSTM PARAMETER FOR SECOND EXPERIMENTAL OUTCOMES

Parameters	Values
Model	Sequential
Hidden layer	4
Hidden unites	50
Neurons per hidden layer	30,30,30,1
Type of layer	LSTM
Activation hidden layer	Softmax
Epoch	100
Optimizer	ALO
Time step	1

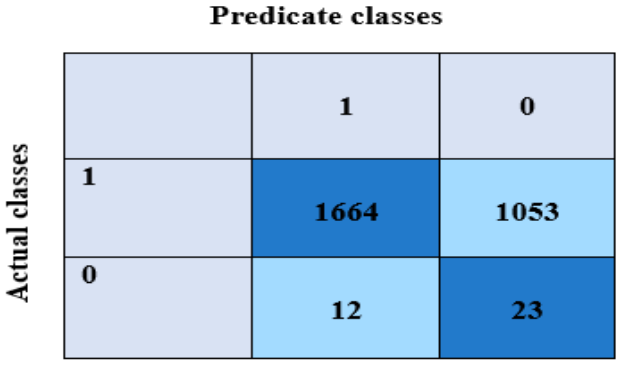


Fig. 8. Confusion matric of LSTM with ALO optimizer.

TABLE IV. COMPARATIVE CLASSIFICATION PERFORMANCES OF LSTM BASED MODEL

Models	Accuracy	Precision	Recall	F1 - score
LSTM based model	91.4%	94.1%	92.1%	93.1%
LSTM with ALO algorithm	98.6%	99.2%	89.6%	98.6%

Based on the findings, we conclude that the model performs well in every experiment, with high precision values ranging from 94.1% to 99.2%. This means that a high rate of the predicted positive instances is correct. The values of recall also show promising performance, ranging from 99.2% to 89.6%, meaning that a substantial proportion of the actual positive samples are predicted correctly. We establish the F1-scores at 98.9%, indicating a good overall performance and maintaining a balance between recall and accuracy. In the second experiment, the ALO algorithm produced an average accuracy, recall, and F1-score of 98.6%, indicating consistently strong classification performance across all classes. This indicates that the classification model is successful in correctly classifying the various thyroid disease classes.

From Table III, the recommended strategy, which selected the optimal hyperparameter for the LSTM using the AlO algorithm, improved the classifier overall by 7.2% compared to the achieved accuracy. With excellent accuracy and consistency across all criteria, the classification performance results show that the model performs robustly and consistently when it comes to differentiating between the different thyroid classes.

As mentioned in references [21], and [22], pre-processing and transforming the study's database into a regular form enhanced the performance of the proposed model. The use of the feature selection algorithm also helped to reduce the number of features used in training the model by focusing on the features with high impact and importance in the prediction process.

Additionally, after comparing the outcomes of the proposed strategy with those of alternative approaches found in the literature. It validates that, even with the same database but a different technique, the efficiency attained by the suggested system for prediction when using the ALO algorithm to enhance the performance of the LSTM algorithm surpasses other research with greater classification accuracy. It was also observed that although some previous research not employed any approaches in feature selection and these factors might have a big impact on how well the classification methods work.

VI. CONCLUSION

The primary goal of this research is to create a new model that predicts thyroid illness by combining the ALO metaheuristic algorithm with the LSTM algorithm. The proposed methodology is based on three steps: preprocessing, feature selection, and classification. By using the commonly used data on thyroid disorders, we trained and evaluated the proposed technique. We employed several metrics, such as accuracy, precision, recall, and F1 score, to assess the model's performance. The results of the experiments showed that 1) using entropy in the feature selection step can cut the total number of features from 31 to 10; and 2) using the ALO algorithm in the suggested strategy made the LSTM classifier 7.2% better overall and gave the best accuracy of 98.6%. The model's high level of accuracy and usefulness illustrate its practical relevance and significance in enhancing patient care. Future studies in this area should concentrate on investigating more intricate model architectures that are capable of capturing even more minute patterns in the thyroid data. Additionally, the model's performance might be improved, and more thorough insights could be obtained by employing different feature selection approaches other than the entropy technique used in this work. We can bridge these gaps and strive for innovative advancements. We can continue to refine the classification models for thyroid illnesses, enabling more accurate and reliable identification of thyroid diseases. This could revolutionize healthcare delivery and improve outcomes for individuals suffering from thyroid conditions.

ACKNOWLEDGMENT

We gratefully acknowledge University of Petra for their invaluable support and resources throughout this research. The realization of our project was significantly aided by our research group and all its members. Their unwavering support and the conducive academic environment provided invaluable assistance throughout this endeavour. Our heartfelt thanks to our colleagues and students for their insightful contributions and steadfast support.

REFERENCES

[1] V. Leso, I. Vetrani, L. De Cicco, A. Cardelia, L. Fontana, G. Buonocore, & I. Iavicoli, " The impact of thyroid diseases on the working life of patients: a systematic review," *International Journal of Environmental Research and Public Health*, vol. 17, no.12, pp. 4295, 2020. https://doi.org/10.3390/ijerph17124295

[2] L. Hegedüs, A. C. Bianco, J. Jonklaas, S. H. Pearce, A. P. Weetman, & P. Perros, " Primary hypothyroidism and quality of life," *Nature Reviews Endocrinology*, vol. 18, no. 4, pp. 230-242, 2022. https://doi.org/10.1038%2Fs41574-021-00625-8

[3] D. Dahiwade, G. Patle, & E. Meshram, "Designing disease prediction model using machine learning approach," *In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1211-1215, 2019. https://doi.org/10.1109/ICCMC.2019.8819782

[4] E. Sonuç, " Thyroid disease classification using machine learning algorithms," *In Journal of Physics: Conference Series*, vol. 1963, no. 1, p. 012140, 2021. https://doi.org/ 10.1088/1742-6596/1963/1/012140.

[5] H. Abbad Ur Rehman, C. Y. Lin, & Z. Mushtaq, "Effective K-nearest neighbor algorithms performance analysis of thyroid disease," *Journal of the Chinese Institute of Engineers*, vol. 44, no.1, pp. 77-87, 2021. https://doi.org/10.1080/02533839.2020.1831967.

[6] F. Temurtas, "A comparative study on thyroid disease diagnosis using neural networks." *Expert Systems with Applications*," vol. 36, no. 1, pp. 944-949, 2009. https://doi.org/10.1016/j.eswa.2007.10.010

[7] W. Ahmad, A. Ahmad, C. Lu, B. A. Khoso, & L. Huang, "A novel hybrid decision support system for thyroid disease forecasting." *Soft Computing*, vol. *22*, pp. 5377-5383, 2018. https://doi.org/10.1007/s00500-018-3045-9

[8] S. Mirjalili, " The ant lion optimizer, "*Advances in engineering software*, vol. *83*, pp. 80-98, 2015. https://doi.org/10.1016/j.advengsoft.2015.01.010

[9] L. Abualigah, M. Shehab, M. Alshinwan, S. Mirjalili, & M. A. Elaziz, "Ant lion optimizer: a comprehensive survey of its variants and applications," *Archives of Computational Methods in Engineering*, vol. *28*, pp. 1397-1416, 2021. https://doi.org/10.1007/s11831-020-09420-6

[10] E. Emary, H. M. Zawbaa, & A. E. Hassanien, "Binary ant lion approaches for feature selection," *Neurocomputing*, vol. 213, pp. 54-65, 2016. https://doi.org/10.1016/j.neucom.2015.06.083

[11] A. Graves, & A. Graves, "Long short-term memory," *Supervised sequence labelling with recurrent neural networks*, pp. 37-45, 2012. https://doi.org/10.1007/978-3-642-24797-2_4

[12] S. Hochreiter, & J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735-1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735

[13] G. Van Houdt, C. Mosquera, & G. Nápoles, " A review on the long short-term memory model, " *Artificial Intelligence Review*, vol. *53 no.*8, pp. 5929-5955, 2020. https://doi.org/10.1007/s10462-020-09838-1

[14] Y. Eren, & İ. Küçükdemiral, "A comprehensive review on deep learning approaches for short-term load forecasting," *Renewable and Sustainable Energy Reviews*, vol. *189*, pp. 114031, 2024. https://doi.org/10.1016/j.rser.2023.114031

[15] S. M. Al-Selwi, M. F. Hassan, S. J. Abdulkadir, A. Muneer, E.H. Sumiea, A. Alqushaibi, & M. G. Ragab, "RNN-LSTM: From applications to modeling techniques and beyond—Systematic review," *Journal of King Saud University-Computer and Information Sciences*, PP. 102068, 2024. https://doi.org/10.1016/j.jksuci.2024.102068

[16] C. Avci, B. Tekinerdogan, & C. Catal, " Analyzing the performance of long short-term memory architectures for malware detection models," *Concurrency and Computation: Practice and Experience*, vol. 35, no. 6, pp. 1-1, 2023. https://doi.org/10.1002/cpe.7581

[17] https://www.kaggle.com/datasets/emmanuelfwerr/thyroid-disease-data.

[18] P. Dhawas, m. a. Ramteke, A. Thakur, P. V. Polshetwar, R. V. Salunkhe, & D. Bhagat, "Big Data Analysis Techniques: Data Preprocessing Techniques, Data Mining Techniques, Machine Learning Algorithm, Visualization," *In Big Data Analytics Techniques for Market Intelligence*, pp. 183-208, 2024. https://doi.org/10.4018/979-8-3693-0413-6.ch006

[19] P. Rodríguez, M. A. Bautista, J. Gonzalez, S. & Escalera, " Beyond one-hot encoding: Lower dimensional target embeddin, " *Image and Vision Computing*, vol. 75, pp. 21-31, 2018. https://doi.org/10.48550/arXiv.1806.10805

[20] A. E. Karrar, "The effect of using data pre-processing by imputations in handling missing values," *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, vol. 10, no. 2, pp. 375-384, 2022. http://dx.doi.org/10.52549/ijeei.v10i2.3730

[21] S. G. O. P. A. L Patro., & K. K. Sahu, "Normalization: A preprocessing stage," *arXiv preprint arXiv:1503.06462*. https://doi.org/10.48550/arXiv.1503.06462

[22] L. Al Shalabi, Z. Shaaban, & B. Kasasbeh, " Data mining: A preprocessing engine, " *Journal of Computer Science*, vol. *2, no.* 9, pp. 735-739, 2006. https://doi.org/10.3844/JCSSP.2006.735.739

[23] X. Song, Y. Zhang, W. Zhang, C. He, Y. Hu, J. Wang, & D. Gong, " Evolutionary computation for feature selection in classification: A comprehensive survey of solutions, applications and challenges," *Swarm and Evolutionary Computation*, vol. 90, pp. 101661, 2024. https://doi.org/10.1016/j.swevo.2024.101661

[24] D. Theng, & K. K. Bhoyar, "Feature selection techniques for machine learning: a survey of more than two decades of research, " *Knowledge and Information Systems*, vol. 66, no. 3, pp. 1575-1637, 2024. https://doi.org/10.1007/s10115-023-02010-5

[25] Priyanka, & D. Kumar, "Decision tree classifier: a detailed survey," *International Journal of Information and Decision Sciences*, vol. 12, no. 3,pp. 246-269, 2020. https://doi.org/10.1504/ijids.2020.10029122

[26] Y. Zhu, D. Tian & F. Yan, F, "Effectiveness of entropy weight method in decision-making, "*Mathematical Problems in Engineering*, vol. 1, pp. 3564835, 2020. https://doi.org/10.1155/2020/3564835

[27] H. Li, F. He, Y. Chen, & Y. Pan," MLFS-CCDE: multi-objective large-scale feature selection by cooperative coevolutionary differential evolution," *Memetic Computing*, vol. 13, pp. 1-18, 2021. https://doi.org/10.1007/s12293-021-00328-7

[28] V. K. Pathak, S. Gangwar, R. Singh, A. K. Srivastava, & M. Dikshit, M. " A comprehensive survey on the ant lion optimiser, variants and applications," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 36, no.4, pp. 511-562, 2024. https://doi.org/10.1080/0952813X.2022.2093409

[29] A. C. Johnvictor, V. Durgamahanthi, R. M. Pariti Venkata, & N. Jethi, " Critical review of bio-inspired optimization techniques. Wiley Interdisciplinary Reviews," *Computational Statistics*, vol. 14, no. 1, pp. e1528, 2022. https://doi.org/10.1002/wics.1528

[30] S. Mirjalili, S. " The ant lion optimizer," *Advances in engineering software*, vol. 83, pp. 80-98, 2015. https://doi.org/10.1007/978-3-030-12127-3_3

[31] S. Kumar, & A. Kumar, "A brief review on antlion optimization algorithm," *In 2018 international conference on advances in computing, communication control and networking (ICACCCN)*, pp. 236-240, 2018. https://doi.org/10.1109/ICACCCN.2018.8748862

[32] A. A. Heidari, H. Faris, S. Mirjalili, I. Aljarah, & M. Mafarja, "Ant lion optimizer: theory, literature review, and application in multi-layer perceptron neural networks," *Nature-inspired optimizers: theories, literature reviews and applications*, PP. 23-46, 2020. https://doi.org/10.1007/978-3-030-12127-3_3

[33] M. Heydarian, T. E. Doyle, & R. Samavi, " MLCM: Multi-label confusion matrix," *IEEE Access*, vol. 10, pp. 19083-19095, 2022. http://dx.doi.org/10.1109/ACCESS.2022.3151048

Balancing Privacy and Performance: Exploring Encryption and Quantization in Content-Based Image Retrieval Systems

Mohamed Jafar Sadik, Dr. Noor Azah Samsudin, Dr. Ezak Fadzrin Bin Ahmad
Faculty of Computer Science and Information Technology (FSKTM),
Universiti Tun Hussein Onn Malaysia (UTHM), Batu Pahat, Johor 86400, Malaysia

***Abstract*—This paper presents three significant contributions to the field of privacy-preserving Content-Based Image Retrieval (CBIR) systems for medical imaging. First, we introduce a novel framework that integrates VGG-16 Convolutional Neural Network with a multi-tiered encryption scheme specifically designed for medical image security. Second, we propose an innovative approach to model optimization through three distinct quantization methods (max, 99% percentile, and KL divergence), which significantly reduces computational overhead while maintaining retrieval accuracy. Third, we provide comprehensive empirical evidence demonstrating the framework's effectiveness across multiple medical imaging modalities, achieving 94.6% accuracy with 99% percentile quantization while maintaining privacy through encryption. Our experimental results, conducted on a dataset of 1,200 medical images across three anatomical categories (lung, brain, and bone), show that our approach successfully balances the competing demands of privacy preservation, computational efficiency, and retrieval accuracy. This work represents a significant advancement in making secure CBIR systems practically deployable in resource-constrained healthcare environments.**

Keywords*—*Content-Based Image Retrieval (CBIR); Convolutional Neural Networks (CNN): Encrypted data; Feature extraction; Fully Homomorphic Encryption (FHE); medical imaging; privacy; quantization; retrieval accuracy

I. INTRODUCTION

Content-Based Image Retrieval (CBIR) systems play a vital role in managing digital images, as they support the sorting and retrieval of images based on their specific content like colour patterns, texture elements or shape structure beyond the external tags or descriptions. By eliminating manual annotation individuality and its limitations, this approach significantly outperforms traditional keyword-based retrieval methods by automating the task. By extracting features and indexing images according to those extracted data, CBIR systems can deliver more intuitive and reliable outcomes by predicating directly from the visual content of an image [1–3].

Incorporating encryption into CBIR systems introduces Encryption is being considered as an essential security element for CBIR systems because of the higher and severe requirements with a larger number of digital images to be protected from unauthorized access. By using encryption, sensitive information within images stays safe from unauthorized access, while still enabling the CBIR system to retrieve the required images. This is especially relevant in contexts where privacy is paramount and the image data is too sensitive to leave unencrypted [4–8].

The medical sector stands as a prominent example where the integration of encryption with CBIR is not just beneficial but essential. Medical images contain private patient information, and their retrieval requires the utmost care to maintain confidentiality. As such, medical image retrieval systems must adapt to perform effectively on encrypted data, ensuring that patient privacy is maintained without compromising the diagnostic value of the images [6, 9,10].

Investigating the technical side, the effectiveness of CBIR can be enhanced with the use of advanced machine learning models such as transformers and Convolutional Neural Networks. These pre-trained models have shown exceptional ability in feature recognition and extraction from large datasets, making them irreplaceable in encrypted image analysis. Transformers handle data sequence well, which can be critical in understanding image context, while Convolutional Neural Networks excel in pattern recognition due to their layered architectural design [11–16].

Quantization further refines the functionality of Convolutional Neural Networks in CBIR by compressing the models without significant loss of performance. This process reduces the demand on computational resources and is particularly advantageous for deploying sophisticated CBIR in environments with hardware limitations or where swift image retrieval is needed [17–20].

Despite the significant advances in Content-Based Image Retrieval systems, particularly in the realm of medical imagery, there remains a critical gap in the development of a generalized framework that seamlessly integrates robust encryption protocols. Current CBIR systems often face a trade-off between encryption strength and retrieval accuracy, with many falling short in providing a secure yet efficient retrieval process that caters to the highly sensitive nature of medical data [15,19, 21–23]. Additionally, the application of such systems is frequently constrained by the computational power required for processing and retrieving high-resolution medical images, which further complicates their deployment in resource-limited settings commonly found in healthcare environments.

Addressing these challenges, our study aims to propose a comprehensive framework for CBIR tailored to the medical domain. The goal is to craft a lightweight, secure encryption

algorithm that adequately protects patient confidentiality without affecting the system's ability to accurately retrieve and analyze medical images. Encryption alone, however, is not the primary focusing point of this framework. The proposed system also seeks to harness the power of Convolutional Neural Network-based pre-trained models, particularly the VGG-16 architecture well-known for its ability in image recognition tasks, to facilitate the feature extraction process in encrypted domains.

Furthermore, to tackle the issue of computational Furthermore, to tackle the issue of computational efficiency, the integration of a quantization approach is crucial. In order to deploy the vgg-16 model into practical usage, it is necessary that its memory footprint and computational needs are significantly reduced thereby making the solution suitable to be deployed even in work environment having low check for computational power. Our use case is under the same limit, by processing through quantization we hope to not only improve efficiency of our system but still keep a high retrieval performance which has been hard trade-off in many existing CBIR systems.

II. Contributions

There are three major contributions in this work to-wards medical CBIR:

A. Developing an Efficiency-Aware Framework

We illustrate a CBIR system for medical images, focusing on computational efficiency and accuracy as the main goals. The system is built for high-throughput needs of medical diagnostics with the modest computational resources common in healthcare settings.

B. Pre-Trained CNN Model

The cornerstone of our retrieval pipeline around which all other piece's center, the VGG-16 model is a pre-trained Convolutional Neural Network we use for feature extraction. In using this well-developed model, our framework can take advantage of its deep learning power to improve the accuracy in medical image analysis.

C. Advanced Approach to Quantization

The framework uses a variety of strategies to quantization, including max quantization, 99% percentile quantization, and Kullback-Leibler Divergence (KL) based quantization. These methods reduce the model's memory footprint and computational load, which is crucial for scalable and efficient CBIR deployment.

III. Literature Survey

A. Feature Extraction and Encryption Methods for CBIR

In the rapidly evolving field of cloud-based privacy-preserving image retrieval, researchers have embarked on a journey to balance the confidentiality of encrypted images with the necessity of efficient retrieval mechanisms. This literature survey begins with foundational encryption techniques that pave the way for more advanced retrieval methods, highlighting the journey from basic encryption to sophisticated integrations of cryptographic algorithms and machine learning models. It provides a comprehensive over-view of the state-of-the-art methods developed to ad-dress the dual challenges of maintaining privacy and ensuring practical utility in cloud environments.

Starting with encryption strategies for image security, works such as [5, 24, 25] introduce novel schemes focusing on the variation of Discrete Cosine Transform (DCT) coefficients, encryption of DC and AC coefficients using stream cipher and scrambling encryption, and the secure retrieval of images in the YUV color space, respectively. These foundational methods set the stage for the development of more complex retrieval mechanisms, emphasizing the need for encrypted images to be both secure and retrievable.

Building on these encryption methodologies, studies [25–29] delve into advanced retrieval mechanisms that operate within the constraints of encrypted do-mains. Huffman-code based retrieval, secure Local Binary Pattern (LBP) features, and the application of Markov processes exemplify the innovative approaches taken to extract meaningful information from encrypted images. These mechanisms showcase the progression towards retrieval systems that are not only secure but also capable of accurately identifying images similar to a query image.

Application-specific solutions, as discussed in works [30–33], further demonstrate the usefulness and practical implications of privacy-preserving image retrieval technologies. The introduction of privacy-preserving Scale-Invariant Feature Transform (SIFT), Convolutional Neural Network (CNN) frame-works for medical data, and systems like PIC (Privacy-preserving Image search system on Cloud) high-light the field's move towards addressing specific needs such as medical diagnosis aid and large-scale image search with fine-grained access control. These tailored approaches underline the importance of developing encrypted image retrieval systems that accommodate to the unique requirements of different domains, emphasizing the critical role of privacy in sensitive applications.

Lastly, comparative analyses and the proposal of new frameworks, as seen in [34–42], reflect the ongoing evolution and refinement of privacy-preserving image retrieval methods. By comparing homomorphic encryption-based techniques with feature/index randomization-based techniques and introducing novel frameworks like IES-CBIR for outsourced storage and retrieval, these works contribute to a deeper under-standing of the trade-offs between security, efficiency, and usability. The development of dynamic verifiable retrieval schemes and multi-indexed hashing approaches represents the cutting edge of research, aimed at improving the precision and security of image retrieval in cloud-based systems. We present a comparative analysis between each of the mentioned methods in Table I. It is found that non-of the existing methods have used VGG16 for encrypted CBIR. Furthermore, to the best of our knowledge, none of the existing works on Content-Based Image Retrieval (CBIR) have optimized the trained model through post-training quantization. Quantization techniques are pivotal for augmenting the computational efficiency of CBIR systems. By diminishing the bit-precision of neural network parameters, quantization paves the way for swifter and more energy-efficient computations across extensive hardware accelerators. This precision reduction not only accelerates computational speed but

also aids in energy conservation, rendering it an indispensable strategy for the deployment of large-scale CBIR systems.

As presented in Fig. 1, the general taxonomy of existing works in the field of Content-Based Image Retrieval (CBIR) can be categorized into four main areas, each focusing on different aspects and method-ologies to enhance privacy-preserving image retrieval. The first category, Encryption Methods for Image Retrieval, aims to develop and apply encryption techniques to secure image data while maintaining its retrievability. Approaches in this category include symmetric and asymmetric encryption schemes, homomorphic encryption for computation on encrypted images, and watermarking techniques for securely embedding retrieval information.

TABLE I. SUMMARY OF METHODS AND FEATURES IN ENCRYPTED CONTENT-BASED IMAGE RETRIEVAL (CBIR) AND RELATED APPLICATIONS

Work	Method of Encryption	Type of Features Extracted	Retrieval Model	Quantization
[24]	Permutation of DCT coefficients	Histogram at each frequency position	Unsupervised/Supervised retrieval using integrated distances and conditional probabilities	×
[25]	Stream cipher and scrambling encryption of DC and AC coefficients	AC coefficients histogram	Statistical comparison of histograms to return closest encrypted images	×
[27]	Stream cipher, permutation cipher for Huffman code	Encrypted Huffman-code histograms	Feature comparison to return similar content images	×
[29]	Color value substitution, block permutation, intra-block pixel permutation	Normalized histogram of encrypted visual words (BOEW model)	Direct similarity measurement between feature vectors on the cloud server side	×
[28]	Big-block permutation, polyalphabetic cipher	Secure Local Binary Pattern (LBP) features	Retrieval based on order-preserving encryption and secure LBP features	×
[43]	Stream cipher and permutation encryption for DCT coefficients	Transition probability matrices (Markov process), SVM classification	Feature extraction and classification to evaluate similarity between encrypted images	×
[30]	Homomorphic encryption	Privacy-preserving SIFT features	Secure SIFT feature extraction and representation in the encrypted domain	×
[32]	Homomorphic encryption	Wavelet-based image features	Secure CBIR for diagnosis aid systems with data confidentiality preservation	×
[33]	Not explicitly mentioned (uses encrypted images)	Features extracted by CNN	Privacy-preserving classification and retrieval using CNN framework	×
[40]	Homomorphic encryption vs. feature/index randomization	Not explicitly mentioned	Comparative analysis for confidentiality-preserving image search	×
[34]	Randomized binary encoding and Gaussian random matrix	Encrypted visual words	Secure index construction for large-scale image retrieval without decryption	×
[42]	Comparable encryption	Not explicitly mentioned	Encrypted image search scheme balancing efficiency and privacy	×
[37]	Feature descriptors extracted by CNN models, encrypted hierarchical index tree	Features extracted by CNN, hierarchical index tree	Similarity search for encrypted images using SEI with enhanced key privacy	×
[38]	Pre-trained CNN model, encrypted index based on K-means clustering	Features extracted by CNN	Dynamic verifiable retrieval over encrypted images (DVREI) scheme	×
[36]	ViT model and ITQ method for multi-indexed hashing (MIH)	Features extracted by ViT model, secure Hamming distance protocol	Privacy-preserving content-based image retrieval using MIH	×
[41]	Cryptographic techniques and secure indexing schemes	Not explicitly mentioned	Content-based retrieval over encrypted multimedia databases	×
[35]	Based on IES-CBIR scheme	Not explicitly mentioned	Outsourced privacy-preserving storage and retrieval in large shared image repositories	×
[44]	Encryption of difference matrices of RGB components (EDH-CBIR)	Euclidean distance between feature vectors to measure similarity	Euclidean distance	×
[45]	Permutation of big-blocks and substitution of binary code of DCT coefficients	Local Markov Features and Bag-of-Words Model	Efficient Content-Based Image Retrieval (CBIR) service for image owners with direct feature extraction	×
[46]	Thumbnail-preserving encryption (TPE)	HSV and uniform local binary pattern (ULBP) features	Thumbnail-Preserving Encryption (TPE)	×
[47]	Fully Homomorphic Encryption (FHE), Secure Multi-Party Computation (SMPC), AES, FPE	Local invariant features	CBIR, Attribute-Based Encryption (ABE), Functional Encryption (FE)	×
[48]	Asymmetric scalar-product-preserving encryption (ASPE)	Local invariant features	Invariant Features Selection	×
[49]	Watermark-based encryption	Dominant local patterns, Relative directional edge patterns (RDEP)	Dominant Local Patterns, Watermark Encryption	×
Ours	Multi-Tiered Texture and Color Encryption (MTTCE)	VGG-16 based features	Efficiency aware VGG-16 retrieval model with supporting three quantization modes	√

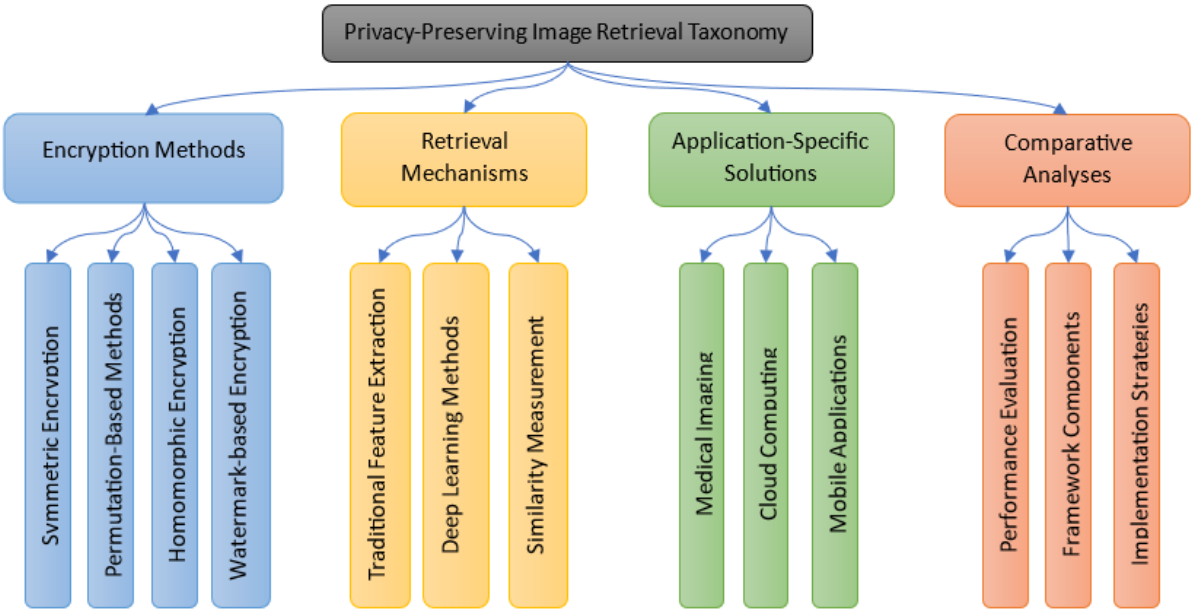


Fig. 1. The general taxonomy of existing works focuses on CBIR.

The second category, Retrieval Mechanisms and Feature Extraction, seeks to enhance the efficiency and accuracy of retrieval mechanisms by optimizing feature extraction and similarity measures. This involves traditional feature extraction methods such as SIFT, SURF, and HOG, as well as deep learning-based feature extraction using convolutional neural networks (CNNs) and similarity measures including Euclidean distance, cosine similarity, and advanced metrics tailored for encrypted data.

The third category focuses on Application-Specific Solutions for Privacy-Preserving Retrieval, developing tailored solutions that address specific application domains and their unique privacy requirements. The fourth and final category, Comparative Analyses and Framework Proposals, involves conducting comparative studies of existing methods and proposing comprehensive frameworks to guide future research and implementation in the field of CBIR. This taxonomy provides a structured overview of the diverse approaches and innovations aimed at enhancing the privacy and effectiveness of image retrieval systems.

B. Efficiency Optimization Methods

The quest for optimizing search efficiency in large-scale image databases has given rise to several innovative strategies in the realm of content-based image retrieval (CBIR). Among these, quantized deep learning frameworks have become a cornerstone for hashing-based retrieval systems. Hashing Nets tackles the challenge of limited storage space and computational resources, especially relevant for satellite remote sensing image retrieval and unmanned aerial vehicles (UAVs). The proposed Quantized Deep Learning to Hash (QDLH) framework introduces binarized weights and activation functions, creating a lightweight neural network that considerably reduces the demand on hardware resources [50]. This need for efficiency extends to other works, such as the application of deep model quantization and compression to CNNs on ASIC chips, achieving comparable performance to floating-point models with a mere 2-bit weights quantization [51].

Addressing the constraints of mobile device capabilities, the OMCBIR framework emerges with an innovative solution. It presents ALNet, an ultra-lightweight neural network that harnesses pointwise group convolution, channel shuffle, and a convolutional attention module, substantially minimizing the model's size without compromising on retrieval accuracy [52].

For the complex domain of remote sensing images, a different approach is introduced in (2021) with the APQ method. This method leverages a multi-scale attention-based CNN combined with an enhanced product quantization technique to efficiently compress features, significantly improving retrieval performance [53].

Furthermore, the domain of multi-label image retrieval presents its own set of challenges, particularly in maintaining semantic integrity within quantization errors. The Multi-label Contrastive Hashing (MCH) technique innovates with a curriculum strategy that carefully adjusts the quantization loss weight, fostering the preservation of multi-level semantic similarity more effectively than prior hashing-based retrieval methods [54].

C. Research Gap

Despite the remarkable progress in content-based image retrieval (CBIR) systems, there exists a visible gap in the domain of medical image retrieval, specifically concerning the post-training optimization of pre-trained models. Current literature prominently features methods optimized for general image databases or remote sensing images, focusing on quantization and hashing to enhance retrieval efficiency and reduce computational demands. These advancements, while substantial, are not directly translatable to the unique requirements and complexity of medical imaging.

Medical images, such as radiographs, MRI, and CT scans, present distinct challenges due to their high-dimensional data and the critical need for precision in feature extraction to capture clinically relevant information. Pre-trained models, often developed on natural image datasets, which may not naturally capture the specific features and diseases that are noticeable in medical images. Fine-tuning pre-trained models on medical datasets is a common strategy to transfer learning from one domain to another; however, there is an obvious absence of research into post-training optimization techniques that refine these models further for the medical domain.

Existing fine-tuning practices largely focus on adapting the pre-trained models to new datasets by retraining some layers while keeping others frozen. This method is beneficial but does not fully exploit the potential of the models to conform to the particularities of medical image analysis. Post-training optimization can involve techniques such as neural architecture search (NAS) tailored to medical datasets, advanced quantization specifically sensitive to medical imaging features, or specialized regularization strategies that address the overfitting risks associated with medical image datasets, which are often smaller and more variable than those used in the training of general models.

The absence of a dedicated post-training optimization phase means that while the models may perform well on general benchmarks, their efficacy in medical scenarios, where the margin for error is minimal, could be significantly enhanced. Such optimizations could lead to improvements in retrieval accuracy, relevance of retrieved images, and ultimately, clinical usefulness. Our article aims to investigate this gap, proposing a framework for post-training optimization of pre-trained models specifically fine-tuned for medical image retrieval, with the goal of maximizing the clinical relevance and accuracy of the retrieved results.

IV. METHODOLOGY

A. Problem Formulation

Given an encrypted image database and a feature extraction model, the system must efficiently process query images to retrieve and return images that are visually similar to the query image. This process involves encryption and decryption of images and features to ensure privacy preservation, requiring the system to address several key challenges:

1) Efficient and secure encryption and decryption mechanisms that maintain the usability and accessibility of the retrieval system.

2) The development of a robust feature extraction model that can effectively operate on encrypted images to extract meaningful features for identification.

3) The implementation of an effective identification and retrieval mechanism that can operate in the encrypted domain, ensuring privacy while maintaining high retrieval accuracy.

4) The development of a quantization of the trained retrieval model to enable efficiency and scalability.

5) The main objective is to achieve a balance between privacy preservation, computational efficiency, and retrieval effectiveness within a cloud-based CBIR system.

More formally, we write the problem based on the following entities and roles:

1) Data owner: The data owner possesses an image database $I = \{i_1, i_2, \dots, i_n\}$ consisting of n images. Each image in the database is encrypted using a unique key from the set $K = \{k_1, k_2, \dots, k_n\}$, resulting in an encrypted image database $E = \{e_1, e_2, \dots, e_n\}$. The data owner also trains a feature extraction model Ψ with E, which is capable of extracting features from encrypted images. The encrypted database E and the model Ψ are then uploaded to the cloud server for storage and deployment.

2) Cloud server: Serving as the backbone for storage and computational power, the cloud server stores the encrypted image database E and the feature extraction model Ψ. Upon receiving an encrypted query image EQ from a query user, the server processes E and EQ through Ψ to extract their features, $F_E = \{f_{e1}, f_{e2}, \dots, f_{en}\}$for the database images and F_{EQ} for the query image. The similarity between F_E and F_{EQ} is assessed using the Euclidean distance metric to identify the k most similar images. The identifiers of these images, $ER = \{er_{ID1}, er_{ID2}, \dots, er_{IDk}\}$, are then returned to the query user.

3) Query user: The query user is interested in retrieving images similar to a query image Q. The query image is first encrypted to EQ and then uploaded to the cloud server. After receiving the encrypted query result set ER from the cloud server, the query user sends these identifiers $ID^R = \{ID_{R1}, ID_{R2}, \dots, ID_{Rk}\}$ back to the data owner to obtain the corresponding decryption keys $RK = \{rk_{IDR1}, rk_{IDR2}, \dots, rk_{IDRk}\}$.

These keys allow the query user to decrypt the received images, resulting in the final retrieval set $R = \{r_{IDR1}, r_{IDR2}, \dots, r_{IDRk}\}$. We present the mathematical symbols used in this article in Table II.

TABLE II. THE SUMMARY OF NOTATIONS

Notations	Definitions
n	The size of the image dataset
m^2	Number of blocks
$I = \{i_1, i_2, \cdots, i_n\}$	The plaintext image dataset
$E = \{e_1, e_2, \cdots, e_n\}$	The encrypted image dataset
$K = \{k_1, k_2, \dots, k_n\}$	The set of security keys
$F_E = \{f_{e1}, f_{e2}, \cdots, f_{en}\}$	The encrypted image feature dataset
Q	The plaintext query image
EQ	The encrypted query image
F_{EQ}	The encrypted query image feature
Ψ	The feature extraction model
$\mathrm{ER} = \{\mathrm{er}_{\mathrm{ID1}}, \mathrm{er}_{\mathrm{ID2}}, \cdots, \mathrm{er}_{\mathrm{IDk}}\}$	The encrypted query result images
$\mathrm{ID}_R = \{\mathrm{ID}_{R1}, \mathrm{ID}_{R2}, \cdots, \mathrm{ID}_{Rk}\}$	The query result image ID
$\mathrm{RK} = \{\mathrm{rk}_{\mathrm{IDR\ 1}}, \mathrm{rk}_{\mathrm{IDR\ 2}}, \cdots, \mathrm{rk}_{\mathrm{IDRk}}\}$	The key corresponding to the resulting image
$R = \{r_{\mathrm{IDR1}}, r_{\mathrm{IDR2}}, \cdots, r_{\mathrm{IDRk}}\}$	The query result image
$\mathrm{sub}\,I = \{\mathrm{sub}\,I_1, \mathrm{sub}\,I_2, \cdots, \mathrm{sub}\,I_{m^2}\}$	Image subblock
$\mathrm{sub}\,E = \{\mathrm{sub}\,E_1, \mathrm{sub}\,E_2, \cdots, \mathrm{sub}\,E_{m^2}\}$	Encrypted image subblock

B. System Architecture

The sequence diagram showed in Fig. 2 presents the operational workflow of a privacy-preserving Content-Based Image Retrieval (CBIR) system, illustrating the interactions among three key entities: the Data Owner, the Cloud Server, and the Query User. The process begins with the Data Owner, who possesses a database of images intended for secure retrieval. This entity takes the initial steps by encrypting the image database, creating an encrypted image database (E), and developing a feature extraction model (Ψ). These components are crucial for ensuring privacy and facilitating feature-based image retrieval in an encrypted domain. Once prepared, the Data Owner uploads both the encrypted image database and the feature extraction model to the Cloud Server, a platform that provides the necessary storage and computational power for the system. Upon the system's readiness to handle queries, the Query User engages by first encrypting a query image (EQ) using a similar encryption methodology as the Data Owner. This encrypted query image is then uploaded to the Cloud Server, indicating the start of the retrieval process. The Cloud Server, leveraging the previously uploaded feature extraction model, processes both encrypted query image and encrypted image database to extract their respective features. Although the specific approach for measuring similarity is abstracted in this diagram, the Cloud Server identifies the k most similar images to the query and returns their identifiers (ER) to the Query User. The interaction between the Query User and the Data Owner is reinitiated when the Query User requests the decryption keys for the received images. The Data Owner responds by providing the necessary decryption keys (RK), enabling the Query User to decrypt and access the final retrieval set of images (R). This sequence diagram effectively encapsulates the secure and private workflow of a CBIR system, emphasizing encryption for privacy, cloud-based feature extraction and retrieval, and decryption for accessing the retrieved images, all while abstracting the complexities of similarity measurement and feature extraction details.

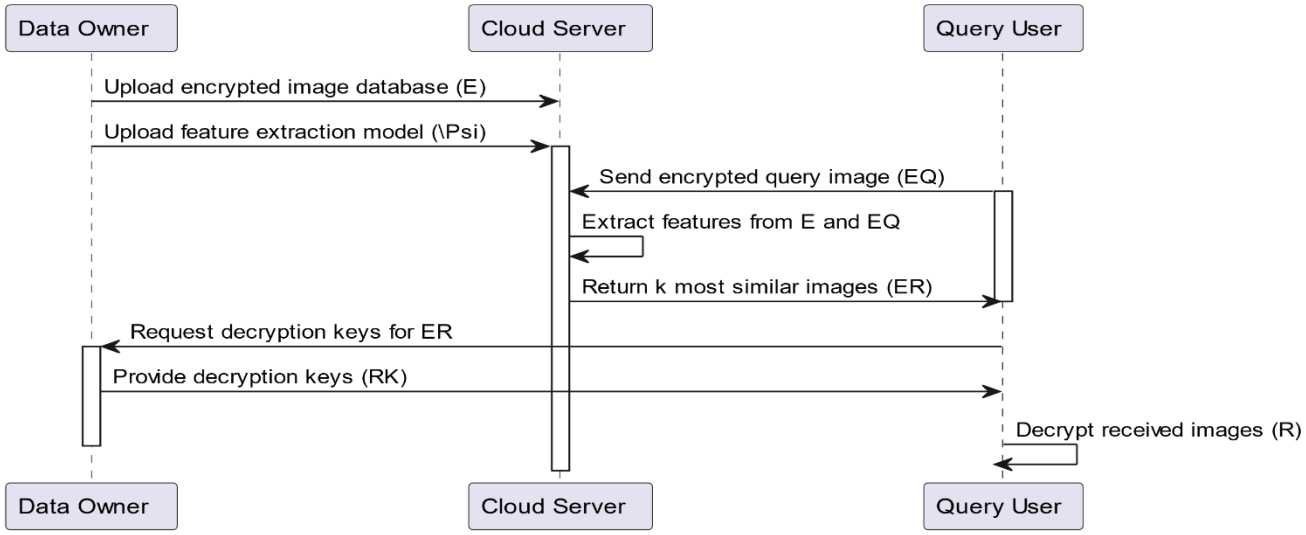


Fig. 2. Sequence diagram illustrating the operational flow of a privacy-preserving Content-Based Image Retrieval (CBIR) system.

C. Framework

The framework proposed in Fig. 3 provides a holistic view of encrypted image retrieval with modified VGG-16 architecture and state-of-the-art quantization methods. It helps in solving the dual issues of ensuring privacy through encryption and enabling efficient content-based image retrieval.

The process started with the encryption of data, all images from the training distribution were then encrypted and stored in cloud. This step makes certain that data privacy is met at every point of retrieval. The basis of the framework is a VGG-16 feature extraction module adjusted to work with encrypted images. Using the encrypted input image, convolutional layers and max pooling is applied on top of this leading to fully connected layers that give an output feature vector.

The same framework also introduces a key model optimization module just to keep the model's memory and efficiency. There are four quantization methods evaluated: Max Quantization, KL Quantization, 99 Percentile Quantization method and Full Model Method. These techniques were applied to the original model weights to obtain a quantized model that trades retrieval accuracy for computational and memory efficiency.

The quantized model is followed by a fine-tuning step on the encrypted dataset preparing its parameters, then to suit the encrypted medical images. The tuned model is then deployed to cloud where given query images while returns the K similar images in database.

The framework also covers secure user interaction flow as well. The users are able to send encrypted query images, receive the encrypted results and ask for decryption keys from data owners to obtain the recognized multimedia files. This means the secure conversion of text throughout its lifecycle as it is fetched.

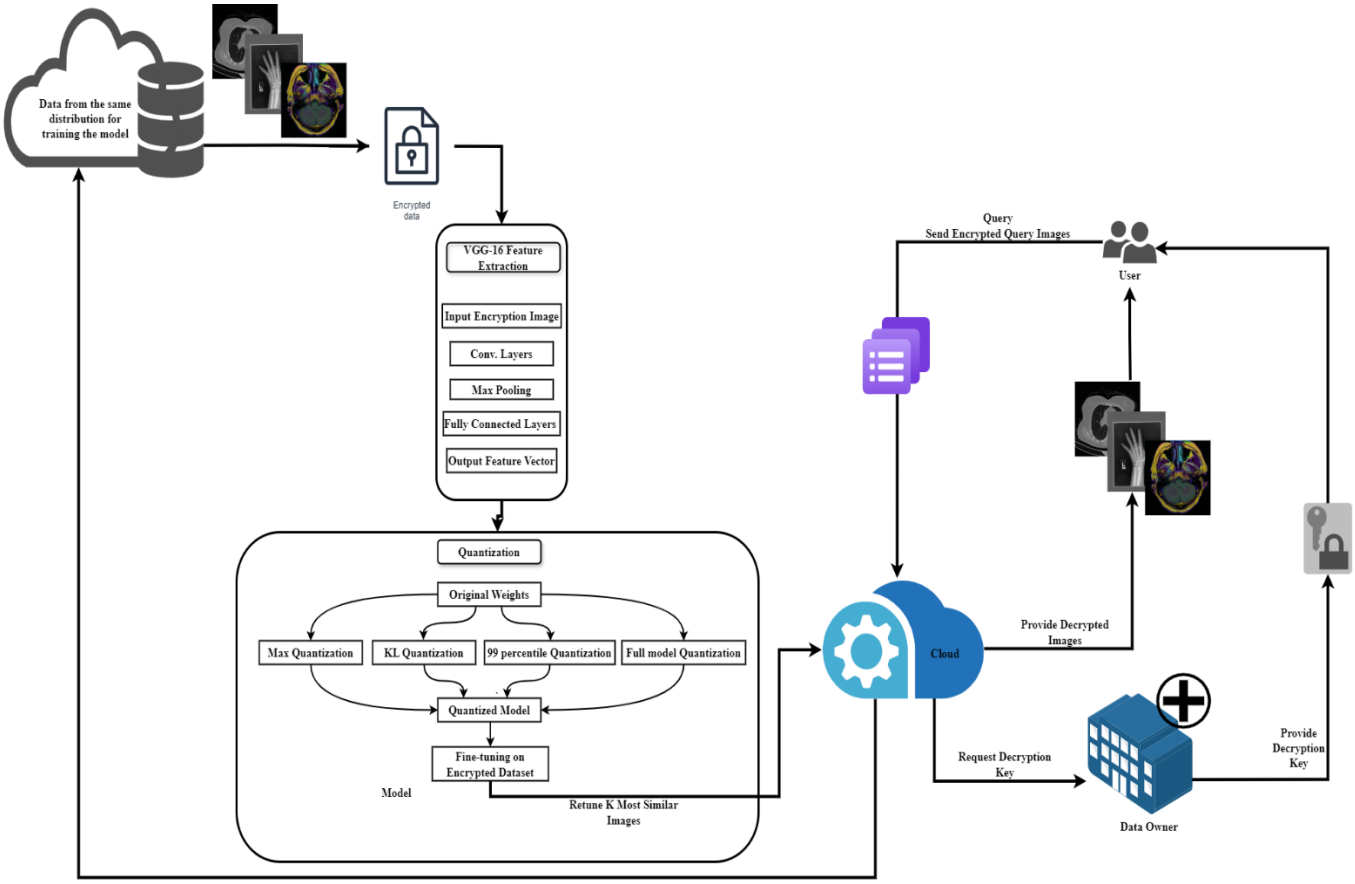


Fig. 3. Encrypted medical image retrieval framework using VGG-16 and quantization techniques.

In addition to support for advanced encryption, feature extraction and quantization as well secure user interactions this framework presents one of the most state-of-the-art tools in privacy-preserving content-based image retrieval especially useful e.g., when it comes to dealing with sensitive domains like medical imaging.

D. Encryption

The encryption algorithm detailed in the provided pseudocode is a comprehensive multi-stage process designed to enhance the security of content-based image retrieval (CBIR) systems. It ensures both local and global image information is encrypted, safeguarding essential retrieval features while preventing unauthorized access. By methodically encrypting local textures, global textures, and color information, this algorithm offers a robust solution to maintaining the privacy and security of image data in CBIR applications. The provided pseudocode outlines a multi-stage encryption algorithm designed to enhance the security of content-based image retrieval (CBIR) systems. The primary objective is to secure both local and global information of an image, ensuring that features necessary for retrieval are preserved while the image remains protected from unauthorized access. The algorithm begins with the input of an original image, denoted as I. The output of the algorithm is the encrypted image I_{enc} and the encryption key K, which is essential for decrypting the image. The start of the process involves dividing the original image I into non-overlapping subblocks, referred to as B_i . This segmentation is crucial as it sets the stage for detailed encryption targeting local textures within the image. In the first stage of encryption, the algorithm focuses on obscuring local texture information within each subblock. For every subblock B_i, the positions of the RGB channel values are scrambled. These scrambling hides local textures, making it difficult for unauthorized users to recognize patterns that could lead to information leakage. Next, the algorithm shifts to protect global texture information. This is accomplished by randomly scrambling the positions between the subblocks B_i. By shuffling these blocks, the spatial relationships within the image are altered, further confusing any attempts to understand the encrypted image's structure. The final stage of encryption deals with securing global color information. For each scrambled subblock B_i, the algorithm substitutes the RGB channel values and swaps the channels. This step not only secures the color information but also ensures that the value substitution is fixedly related to the position of the encrypted subblocks, adding an additional layer of complexity to the encryption scheme. To complete the encryption process, the algorithm generates an encryption key K using a built-in random function. This key is essential for the decryption procedure, allowing authorized users to reverse the encryption process and recover the original image. The processed subblocks are then combined to form the encrypted image $I_{\text{enc.}}$. The encrypted image is now secured, containing no recognizable original content, and can only be decrypted with the correct encryption key. The end of the algorithm marks the completion of the encryption process.

Algorithm 1: Multi-Stage Image Encryption for CBIR Systems

1. **Input:** Original Image I
2. **Output:** Encrypted Image I_{enc}, Encryption Key K
3. Start
4. Divide the original image I into non-overlapping subblocks B_i
5. **for** each subblock B_i **do**
6. Scramble the positions of the RGB channel values within B_i to hide local texture information
7. **end for**
8. Randomly scramble the positions between subblocks B_i to obscure global texture information
9. **for** each scrambled subblock B_i **do**
10. Substitute the RGB channel values and swap the channels to secure global color information
11. **end for**
12. Generate the encryption key K using a built-in random function
13. Combine the processed subblocks to form the encrypted image I_{enc}
14. **End**

It would summarize the encryption algorithm steps well and in a logical, flowing manner, the multistage operation will tend to be secure to CBIR systems more as it will protect the local histogram features and unlink global texture color information. Intrablock scrambling, interblock scrambling and channel substitution are combined to give strong encryption that can be used to provide protection from access to the image data without a key yet enable integrity for the features of useful image retrieval.

As an illustrating example, we present the results of applying the methods to medical images, showcasing a hand X-ray, an MRI brain scan, and a skull X-ray as depicted in Fig. 4. The original images show distinct medical details essential for diagnosis: the hand X-ray reveals the bone structure, the MRI scan distinguishes soft tissue contrasts within the brain, and the skull X-ray clearly outlines the facial bone anatomy. In their encrypted state, these images transform into a mosaic of indistinguishable colored blocks, effectively covering any diagnostic information and ensuring data privacy. The decrypted images, when compared with the originals, show no apparent loss of detail or quality, indicating that the encryption process is fully reversible and maintains the integrity of medical information. This demonstrates the encryption method's potential for securing sensitive health information as well as efficient storage and distribution of medical images are feasible while maintaining their reversibility for medical use. The encryption approach, therefore, strikes a balance between protecting patient confidentiality and preserving the utility of medical images for diagnostic purposes. This highlights the robustness of the encryption method in terms of its practical use in healthcare settings, where patient data protection is essential without limiting healthcare professionals' capability to examine and evaluate medical imaging as needed.

The security analysis of the algorithm is presented in the following steps:

Step 1: Dividing the Image into Subblocks

Let I be the original image of size $M \times N$. The image is divided into non-overlapping subblocks B_i of size $m \times n$, where $m \times n$ is a divisor of $M \times N$. Let k be the number of subblocks, so

$$k = \frac{M \times N}{m \times n} \quad (1)$$

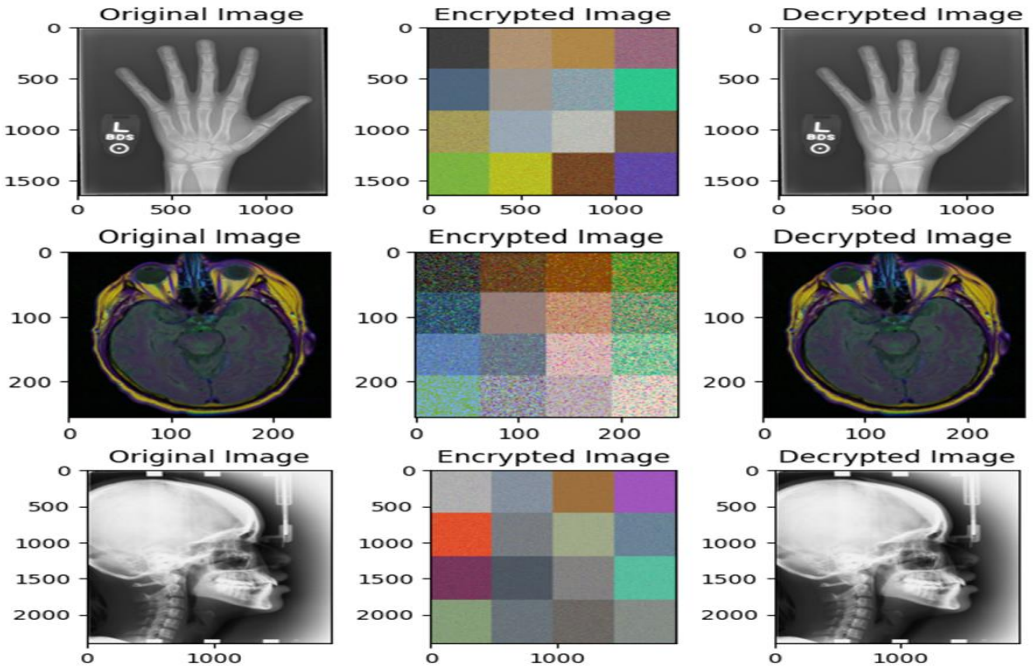


Fig. 4. Encryption process illustrated: original, encrypted, and decrypted states of medical images for secure image retrieval.

Step 2: Scrambling the RGB Values Within Each Subblock

For each subblock B_i, the positions of the RGB channel values are scrambled. The number of possible permutations for each subblock is $(m \times n)$!. Given k subblocks, the total number of possible permutations is:

$$(m \times n)!^k \quad (2)$$

This large permutation space ensures that local textures are effectively obscured.

Step 3: Scrambling the Positions of Subblocks

After local texture encryption, the subblocks are shuffled. The number of possible permutations of k subblocks is !. Combining this with the permutations from the previous step, the total number of possible configurations is:

$$(m \times n)!^k \times k! \quad (3)$$

This significantly increases the complexity, making it difficult for unauthorized users to reconstruct the original image based on subblock positions.

Step 4: Substituting RGB Values and Swapping Channels

For each subblock, the RGB channel values are substituted, and the channels are swapped. Suppose each RGB value can be replaced with any other value within the range $[0,255]$.

The number of possible substitutions for each channel is 256. Since there are three channels, the total number of possible substitutions for each subblock is:$(256)^3$

Swapping the channels adds an additional 3 ! permutations.

Therefore, the total number of configurations for each subblock considering color information is: Combining all the steps, the total number of possible permutations for the entire image encryption process is:

$$(m \times n)!^k \times k! \times (256)^3 \times 3! \quad (4)$$

This enormous number represents the total permutation space, which is computationally infeasible to brute force.

The encryption key K is generated using a built-in random function. The key must be sufficiently long and complex to cover the permutation space generated by the above steps. Assume the key length is L. The entropy of the key is:

$$(K) = L\log_2(N) \quad (5)$$

where, N is the number of possible values for each part of the key. A sufficiently large L ensures that the key space is large enough to resist brute-force attacks.

The security of this encryption algorithm is primarily based on the vast permutation space created by scrambling subblock positions, substituting RGB values, and swapping channels. The combined permutations create an extremely large key space, making brute-force attacks impractical. The algorithm's design ensures that both local and global features of the image are secured, providing robust protection against unauthorized access while maintaining the integrity necessary for effective CBIR system retrieval.

E. VGG-16

The VGGNet architecture, a highly influential convolutional neural network (CNN), was developed through a partnership between the Visual Geometry Group at the University of Oxford and Google DeepMind. It represents a significant evolution in the CNN landscape, building upon the foundational principles established by its predecessor, AlexNet. VGGNet has become well-known for its architectural depth and the use of uniformly small convolutional filters, specifically 3x3 kernels, which have set a new standard for feature extraction in image recognition tasks. As shown in Fig. 5, the architecture consists of a repeating pattern of convolution layers followed by 2x2 max pooling layers, maintaining the spatial hierarchy of the features being learned. This approach not only maintains the spatial hierarchy of the features being learned but also allows for an increase in the depth of the network without a corresponding explosion in computational complexity. By deepening the network, VGGNet significantly enhances the hierarchical feature learning process, capturing fine-grained details that are often crucial for accurate image classification. The VGGNet family comprises several models, among which the VGGNet-16 and VGGNet-19 are the most notable. These models are differentiated by the number of weight layers they contain: 16 and 19, respectively. Both have demonstrated remarkable performance on large-scale image recognition tasks, contributing to their widespread adoption in the field. The depth of these networks has proven to be a key factor in their ability to perform complex image classifications with high accuracy, making them particularly useful in applications where precision is critical. Furthermore, the impact of VGGNet extends beyond its immediate performance. The architecture has provided invaluable insights into the design of deep neural networks, influencing subsequent innovations in the domain. Its widespread use as a pre-trained model for a variety of tasks points out its significance because it provides a starting point for additional optimizing and adaptation for specialized applications, such as those in medical imaging and other fields where detailed feature detection is crucial [55–57].

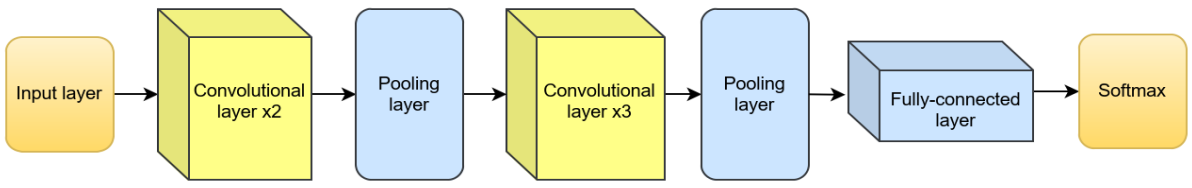


Fig. 5. Architectural flowchart of a deep convolutional neural network with repeating convolution and pooling layers, culminating in a softmax output.

F. Quantization

Content-Based Image Retrieval (CBIR) is a technique for retrieving relevant images from a database based on a given query image. When dealing with encrypted images, this task becomes more complex. To address this, we propose a method using a VGG-16 model with various quantization techniques to improve efficiency and performance. The first step involves applying quantization to the last layer of a pre-trained VGG-16 model. Quantization reduces the precision of the weights, leading to a smaller model size and faster computations. The quantization methods used include max quantization, KL quantization, 99 percentile quantization, and full model quantization. For max quantization, the maximum absolute value of the weights is calculated, and each weight is divided by this value. In KL quantization, the histogram of the weights is computed, and quantization thresholds are determined using Kullback-Leibler divergence, followed by quantizing the weights based on these thresholds. For 99 percentile quantization, the 99th percentile of the absolute weights is calculated, and weights are divided by this value. Full quantization applies quantization to all layers of the model. After quantization, the original weights of the last layer are replaced with the quantized weights, resulting in a quantized model. The quantization pseudocode is presented in Algorithm 2.

Algorithm 2: Quantization for Last Layer

1: **Input**: Trained model M, quantization mode *mode*
2: **Output**: Quantized model M'
3: W← weights of the last layer of M
4: **if** *mode* == "max" **then**
5: $W_{max} \leftarrow \max(|W|)$
6: Quantize W as $W_q = W / W_{max}$
7: **else if** *mode* == "KL" **then**
8: Compute the histogram of W
9: Determine the quantization thresholds using KL divergence
10: Quantize W using the determined thresholds
11: **else if** *mode* == "99%" **then**
12: $W_{99} \leftarrow 99^{th}$ percentile of $|W|$
13: Quantize W as $W_q = W / W_{99}$
14: **else if** *mode* == "fully" **then**
15: Quantize all layers of M
16: **end if**
17: Replace the last layer weights of M with W_q
18: **return** M'

Algorithm 3: Encrypted CBIR with VGG-16 Quantized Models

1: **Input:** Encrypted image dataset D, query image Q, quantization mode *mode*
2: **Output:** Retrieved images
3: Initialization:
4: Load pre-trained VGG-16 model M
5: Apply Quantization for Last Layer algorithm with mode\text{mode}mode
6: Fine-tune the quantized model Mon encrypted images
7: Perform image retrieval using the quantized model M'
8: **return** Retrieved images

G. Big O Notation

Quantization is a technique used to reduce the precision of the weights in a neural network, which in turn reduces the memory footprint and computational requirements of the model. Here's a complexity analysis of the memory usage after applying different quantization methods to the last layer of the VGG-16 model as described in the algorithms.

Let's denote the number of weights in the last layer of the VGG-16 model as W. Typically, these weights are stored as 32-bit floating-point numbers (i.e., each weight takes 4 bytes).

The memory usage for the last layer before quantization is:

$$\text{Μεμορψ}_{\beta\varepsilon\phi o\rho\varepsilon} = W \times 4\beta\psi\tau\varepsilon\sigma \quad (6)$$

Quantization reduces the precision of these weights, typically to 8-bit integers (i.e., each weight takes 1 byte). The memory usage after quantization depends on the number of weights and the quantization method used.

Quantization reduces the precision of these weights, typically to 8-bit integers (i.e., each weight takes 1 byte). The memory usage after quantization depends on the number of weights and the quantization method used.

Max quantization scales the weights based on the maximum absolute value. The quantized weights are stored as 8-bit integers, and an additional scaling factor needs to be stored (usually as a 32-bit float).

The memory usage for max quantization is:

$\text{Memory}_{\max} = W \times 1$ byte +4 bytes (scaling factor)

KL quantization uses histograms and thresholds determined using KL divergence. Similar to max quantization, the weights are stored as 8-bit integers, and additional parameters for thresholds may be stored (assuming negligible storage for histograms and thresholds compared to the number of weights).

H. KL Quantization

KL quantization uses histograms and thresholds determined using KL divergence. Similar to max quantization, the weights are stored as 8-bit integers, and additional parameters for thresholds may be stored (assuming negligible storage for histograms and thresholds compared to the number of weights).

The memory usage for KL quantization is:

$\text{Memory}_{\text{KL}} = W \times 1$ byte +4 bytes (scaling factor)

99 Percentile Quantization

99 percentile quantization scales the weights based on the 99th percentile of their absolute values. Again, weights are stored as 8-bit integers with an additional scaling factor.

The memory usage for 99 percentile quantization is:

$\text{Memory}_{99\%} = W \times 1$ byte +4 bytes (scaling factor)

Full quantization applies to all layers of the model. For simplicity, let's denote the total number of weights in the VGG-16 model as $W_{\text{total.}}$.

The memory usage for full model quantization is:

$$\text{Memory}_{\text{full}} = W_{\text{total}} \times 1 \text{ byte} + \text{scaling factors}$$

If there are L layers in the VGG-16 model, and each layer has a scaling factor, the memory usage for scaling factors is negligible compared to the total number of weights.

The memory usage complexity remains linear in terms of the number of weights, but the actual memory usage is significantly reduced due to the lower precision (8-bit vs. 32-bit). In summary, quantization effectively reduces the memory usage of the model by a factor of 4 (since 32-bit weights are converted to 8-bit weights), while the complexity in terms of the number of weights remains the same. This reduction is crucial for deploying models in resource-constrained environments and for speeding up computations during inference. In summary, quantization effectively reduces the memory usage of the model by a factor of 4 (since 32-bit weights are converted to 8-bit weights), while the complexity in terms of the number of weights remains the same. This reduction is crucial for deploying models in resource-constrained environments and for speeding up computations during inference.

V. Experimental Results and Analysis

The fundamental components and their parametric relationships in the proposed framework are systematically presented in Table III. The framework's architecture comprises three main parameter categories: encryption parameters, VGG-16 model configurations, and quantization methods. The encryption parameters, particularly the block size, represent a crucial design choice that determines the balance between security strength and feature preservation. Similarly, the RGB channel substitution parameter provides flexibility in controlling the degree of visual information protection while maintaining essential image characteristics for retrieval purposes. In the deep learning component, the VGG-16 model parameters - learning rate and batch size - establish the foundation for stable model training and resource utilization. The quantization parameters offer different approaches to model compression, each with its unique characteristics: max quantization prioritizes compression efficiency, 99% percentile quantization aims for balanced preservation of significant features, and KL divergence quantization focuses on maintaining statistical distributions. This systematic organization of parameters provides system designers with a clear framework for making informed decisions based on their specific requirements for privacy preservation, computational efficiency, and retrieval capability.

The study leveraged three distinct datasets, each encompassing 1200 medical images. The datasets were further segregated for training and validation, and testing purposes, employing a 1000:100:100 split, respectively. Notably, the datasets were categorized based on anatomical regions, specifically the lung, brain, and bone. This stratified and anatomically categorized data structure facilitated the robust evaluation of the proposed method across diverse medical imaging domains, we present our experimental design in Table IV.

TABLE III. Experimental Design for Comparing our Proposed Method with Benchmark

Parameter Category	Parameter	Value/Range	Trade-offs	Recommended Setting
Encryption	Block Size	1-8 blocks	- Larger blocks: Better retrieval, lower security - Smaller blocks: Higher security, poor retrieval	8 blocks for medical applications requiring balance of privacy and accuracy
	RGB Channel Substitution	0-255	Security vs. Feature quality	Application-dependent, moderate values (128-192) for balanced performance
VGG-16	Learning Rate	0.001	Training speed vs. Stability	0.001 (demonstrated optimal convergence)
	Batch Size	32	Memory usage vs. Training stability	32 (balances resource usage and stability)
Quantization	Max Quantization	Maximum value in layer	Compression vs. Accuracy	When maximum compression is needed
	99% Percentile	99th percentile value	Moderate compression with optimal accuracy	Default choice for most applications
	KL Divergence	Distribution-based	Computational complexity vs. Precision	When distribution preservation is critical

TABLE IV. Experimental Design for Comparing our Proposed Method with Benchmark

	VIT	VGG16
Experiments	Plaintext- encrypted image where block size (1,8)	Plaintext- encrypted image where block size (1, 8)
Pretrained	TRUE	TRUE
Num Classes	3	3
Optimizer	Adam	Adam
Learning Rate	0.001	0.001
Loss Function	CrossEntropyLoss	CrossEntropyLoss
Epochs	100	100
Batch Size	32	32

A. Encryption Quality Analysis

For the experimental evaluation of the encryption, we compare the encryption algorithm adopted with the widely recognized AES standard as shown in Table V. The comparison is grounded on a set of metrics that typically measure the efficacy of encryption methods in terms of security and the ability to resist statistical analysis.

- Entropy: Our algorithm records an entropy of 7.392, which indicates a substantial amount of randomness, although slightly less than the optimal value of 8. This suggests that while the algorithm introduces randomness, there may be room for improvement. The AES, in an unusual turn, shows an entropy value extremely close to zero, which typically would suggest a lack of randomness; however, this could imply a perfect encryption where the output is indistinguishable from a completely random source.
- MSE (Mean Squared Error): In this evaluation, a higher MSE between the original and encrypted images indicates stronger encryption. The MSE for our algorithm is significantly higher than that of AES, suggesting that our method may provide a more robust alteration of pixel values, thereby potentially increasing security.
- SSIM (Structural Similarity Index): The SSIM for our algorithm is noticeably higher than that of AES, implying that the structural integrity of the image is somewhat retained. In contrast, AES's lower SSIM reinforces its role as a robust encryption standard by substantially altering the image structure to secure the data effectively.
- PSNR (Peak Signal-to-Noise Ratio): The lower PSNR associated with our algorithm complements the high MSE, affirming the extensive alteration from the original image. Meanwhile, AES exhibits a higher PSNR, which is usually indicative of a decryption process that maintains image quality. However, in the encryption phase, a lower PSNR may be more desirable as it indicates a greater level of distortion.

B. Image Retrieval Performance Analysis

The confusion matrices and the classification metrics presented in Fig. 6(a), 6(b), and 6(c) illustrates the performance of a Vision Transformer (ViT) based image retrieval system under different conditions: plaintext images and images encrypted using 8 blocks and 1 block encryption. These matrices provide insight into how encryption affects the model's ability to correctly classify images into one of three categories: bone, chest, and MRI.

TABLE V. COMPARATIVE EVALUATION OF ENCRYPTION METRICS: OUR ALGORITHM VERSUS AES

Metric	Our Algorithm	AES
Entropy	**7.392129887**	1.4426951601859516e-10
MSE	**12766.02**	2625.48
SSIM	0.273204	**0.00283838**
PSNR	**7.070247655**	13.93870288

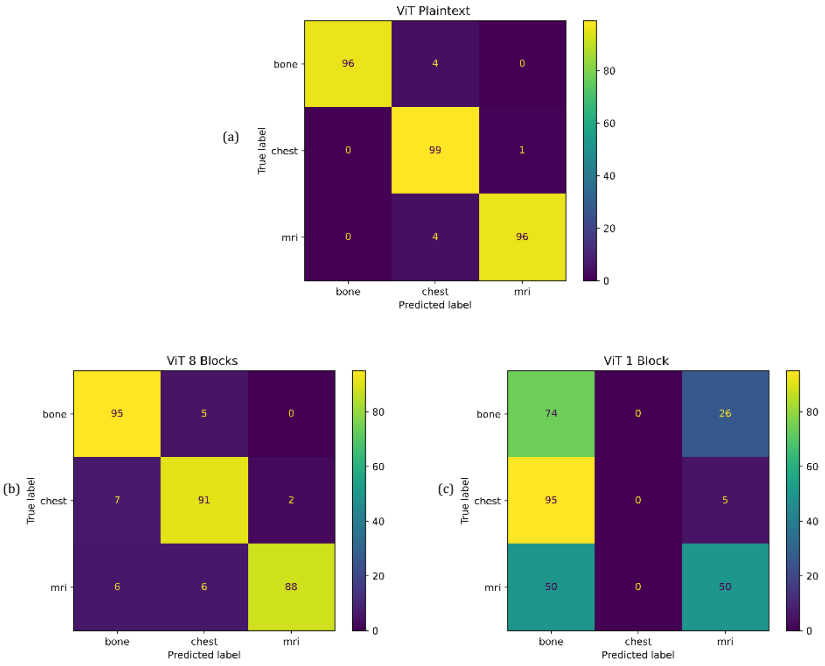


Fig. 6. Confusion matrix of ViT based image retrieval (a) plaintext (b) encrypted domain using 8 block encryption (c) encrypted domain using 1 block encryption.

In the plaintext scenario, the performance of the ViT model is well-exemplified since correctness can be identified with great precision for each category in three classes. It is particularly very high for the classes of bone and MRI, of which only few images can misclassify. The chest category provides a high accuracy impression too but slightly messes up with the remaining categories. A high true positive rate for each class actually represents that the model was well tuned for the features of each class.

The 8 blocks encryption has somewhat degraded performance. It still maintains relatively high accuracy, though an increase in misclassification is seen, especially with the bone and chest categories. A minimal decrease in accuracy is observed in the MRI category. This hints that most probably the model starts having difficulty extracting features when receiving high-encryption images. That is one possible reason for the weakness in performance, considering that encryption puts noise in images, thus blurring features that our model depends on for classification.

With 1 block encryption, the performance drops significantly, most notably in the bone and MRI categories. The model's ability to correctly identify bone images is drastically reduced, and there is a considerable increase in the misclassification of MRI images as bone or chest. The SSIM for this encryption level is likely quite low, indicating that the structural integrity of the images is heavily compromised, and the model is unable to extract meaningful features for accurate classification.

The gradual performance decline across the matrices from plaintext to 1 block encryption illustrates the trade-off between security and usability in encrypted domains. It underscores the challenge of maintaining feature extraction capabilities for classification tasks while also securing the images against unauthorized access or analysis. This balance is crucial in applications where both the confidentiality of the data and the accuracy of automated systems are of paramount importance.

1) ViT-based image retrieval: Table VI presents overall classification metrics for a Vision Transformer (ViT) based image retrieval system, comparing performance across three different categories (Bone, Chest, MRI) and three models: ViT Plaintext, ViT 8 Blocks, and ViT 1 Block.

ViT Plaintext exhibits exceptional performance with accuracy rates hovering around 98% for all categories. The precision is perfect for 'Bone' and nearly so for 'MRI', indicating almost no false positives for these categories. 'Chest' has a slightly lower precision, which suggests a few more false positives but still maintains a high recall, indicating it successfully identified most true 'Chest' cases.

ViT 8 Blocks shows a drop in performance across all metrics, which is expected as the encryption level increases. Accuracy for 'Bone' and 'Chest' categories decreases by about 4-5%, and for 'MRI' by 3%, compared to the plaintext model. Precision sees a more notable decline, particularly for 'Bone' where it drops by over 12%. This suggests that the 8-block encryption introduces enough ambiguity to affect the model's ability to correctly identify features specific to 'Bone' images.

'MRI' retains high precision but suffers in recall, indicating that while most 'MRI' predictions are correct, the model fails to identify all 'MRI' images, likely due to feature loss in the encryption process. ViT 1 Block shows a significant decrease in performance.

For 'Bone', the accuracy and precision are notably lower, and the recall is moderately high, which may indicate a higher number of false negatives. In 'Chest', precision is undefined, which occurs when the denominator in the precision calculation is zero; this happens if there were no predictions made for the 'Chest' category or all predictions were incorrect. The recall for 'Chest' is 0%, confirming the model did not correctly identify any 'Chest' images. For 'MRI', both precision and recall have decreased significantly, with accuracy being marginally better than 'Bone' but still substantially lower than in the other models.

2) VGG-16-based image retrieval: As provided in Fig. 7(a), 7(b) and 7(c), The confusion matrices and classification results for the VGG-16-based image retrieval present a comprehensive view of the model's performance across three different settings: plaintext, 8-block encryption, and 1-block encryption domains.

TABLE VI. OVERALL CLASSIFICATION METRICS OF VIT BASED IMAGE RETRIEVAL

Model	Category	Accuracy (%)	Precision (%)	Recall (%)
ViT Plaintext	Bone	98.67	100.00	96.00
	Chest	97.00	92.52	99.00
	MRI	98.33	98.97	96.00
ViT 8 Blocks	Bone	94.00	87.96	95.00
	Chest	93.33	89.22	91.00
	MRI	95.33	97.78	88.00
ViT 1 Block	Bone	43.00	33.79	74.00
	Chest	66.67	Undefined	0.00
	MRI	73.00	61.73	50.00

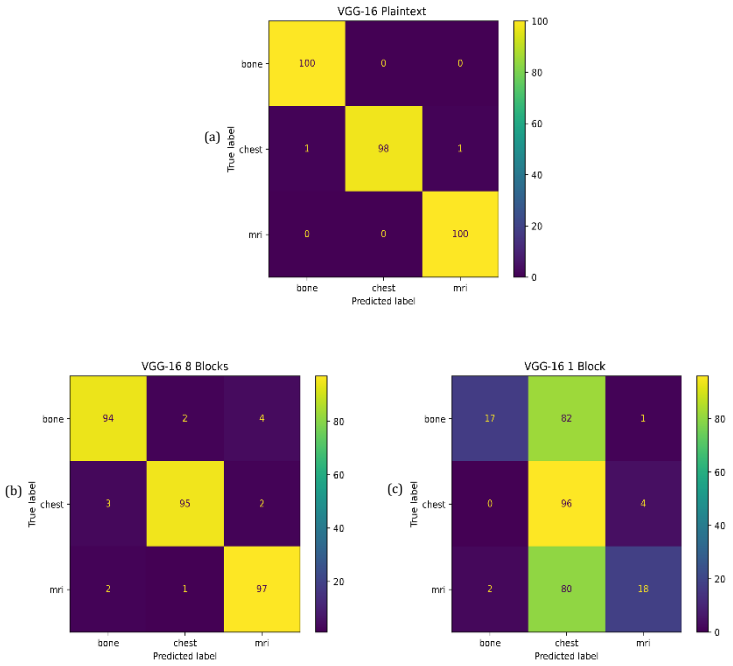


Fig. 7. Confusion matrix of VGG-16 based image retrieval (a) plaintext (b) encrypted domain using 8 block encryption (c) encrypted domain using 1 block encryption.

In the plaintext scenario, the VGG-16 model demonstrates near-perfect classification accuracy, with an impressive 99.33% accuracy across all categories. The precision and recall rates are equally outstanding for the 'Bone' and 'MRI' categories, both hitting 100% in recall, indicating that every relevant image was correctly retrieved. 'Chest' images also show a high level of precision and recall, indicating that the model can distinguish between these medical images with high reliability in an unencrypted domain.

Moving to the 8-block encryption domain, there is a slight but noticeable decrease in performance. The accuracy remains high at 95.33%, yet there are marginal drops in precision and recall for the 'Bone' category, indicating a slight increase in both false positives and false negatives. The Chest and MRI classes lost little accuracy, indicating that encryption adds some ambiguity, however, the model can identify features for retrieval quite effectively.

The 1-block encryption domain delivers quite different findings. The accuracy dropped to 43.67%, indicating that at such high levels of encryption, the model was unable to classify images correctly. Interestingly, the 'Chest' category shows a high recall, suggesting that while the model can identify 'Chest' images, it does so with a high rate of false positives as reflected in the lower precision rate. 'Bone' and 'MRI' categories exhibit poor recall rates, indicating a majority of relevant images are missed, yet when they are identified, they tend to be correct, as shown by the higher precision rates.

Overall, the matrices and results in Table VII highlight the challenges posed by encryption on the ability of CBIR systems to maintain high retrieval performance. They underscore the need for specialized approaches to manage encrypted image data, especially as the strength of encryption increases and significantly impacts the extraction of features critical for accurate image retrieval.

TABLE VII. OVERALL CLASSIFICATION METRICS OF VGG-16 BASED IMAGE RETRIEVAL

Model	Category	Accuracy (%)	Precision (%)	Recall (%)
VGG-16 Plaintext	Bone	99.33	99.01	100.00
	Chest	99.33	100.00	98.00
	MRI	99.33	99.01	100.00
VGG-16 8 Blocks	Bone	95.33	94.95	94.00
	Chest	95.33	96.94	95.00
	MRI	95.33	94.17	97.00
VGG-16 1 Block	Bone	43.67	89.47	17.00
	Chest	43.67	37.21	96.00
	MRI	43.67	78.26	18.00

3) Quantized VGG-16 based image retrieval: In evaluating the performance of the quantized VGG-16 models for image retrieval, we can interpret the provided confusion matrices for each quantization technique as shown in Fig. 8(a), 8(b), 8(c) and 8(d):

a) VGG-16 8 Blocks Final Layer Quantized (Max): The model shows high performance with most 'Bone', 'Chest', and 'MRI' images correctly identified, evidenced by the high numbers on the diagonal of the confusion matrix. There is a slight confusion between the 'Bone' and 'MRI' categories, and to a lesser extent with 'Chest'. However, with 93 out of 100 correct predictions for both 'Bone' and 'Chest' categories, the model demonstrates robustness under the max quantization method.

b) VGG-16 8 Blocks Final Layer Quantized (KL): The model performance is slightly reduced under KL quantization, particularly in the 'Chest' category, where we see a decrease to 91 correct predictions and an increase in misclassification as 'MRI'. The 'MRI' category also shows a minor decrease in performance, with 95 correct predictions. This suggests that the KL quantization may introduce some information loss that affects the model's classification ability.

c) VGG-16 8 Blocks Final Layer Quantized (99%): Under the 99% percentile quantization, the model's performance improves, with 'Bone' and 'MRI' classifications seeing an uptick in correct predictions to 95, and 'Chest' holding steady at 93. This indicates that the 99% percentile method effectively maintains crucial information for image classification.

d) VGG-16 8 Blocks Fully Quantized: The fully quantized model reveals a drastic change in the confusion pattern, with most 'Bone' and 'Chest' images being misclassified as 'MRI'. While this may appear to be a major decline in model performance, it is worth noting that such a high degree of quantization is likely to significantly reduce the model's size and computational requirements. This trade-off between size and accuracy may be beneficial in specific applications where computational efficiency is prioritized over classification accuracy.

Table VIII presents classification metrics for a VGG-16 model with the final layer quantized under different schemes: maximum, KL divergence, 99%, and fully quantized, across medical image categories (Bone, Chest, MRI). The performance metrics detailed—accuracy, precision, and recall—provide insights into the impact of these quantization methods on the model's ability to effectively process encrypted medical images segmented into 8 blocks.

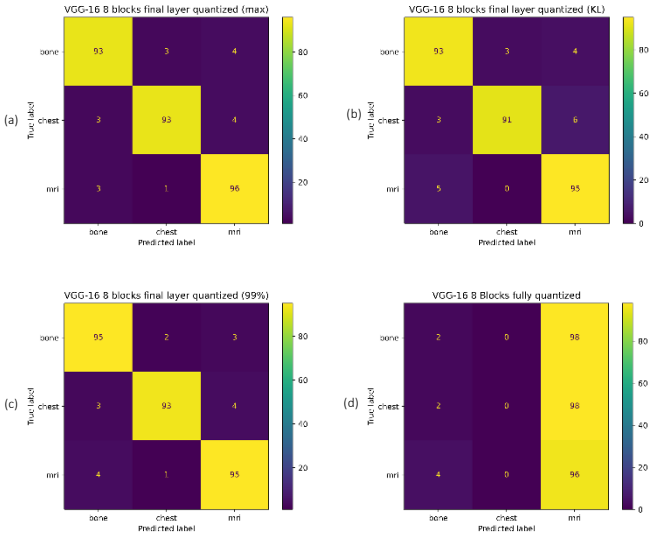


Fig. 8. Confusion matrix of Quantized VGG-16 based image retrieval (a) max (b) KL encryption (c) 99% percentile (d) fully quantized.

TABLE VIII. CLASSIFICATION METRICS OF QUANTIZED VGG-16 FOR DIFFERENT FINAL LAYER QUANTIZATION TYPES MAXIMUM, KL, 99%, AND FULLY QUANTIZED

Model	Category	Accuracy (%)	Precision (%)	Recall (%)
VGG-16 8 blocks final layer quantized (max)	Bone	0.94	0.939	0.93
	Chest	0.94	0.958	0.93
	MRI	0.94	0.923	0.96
VGG-16 8 blocks final layer quantized (KL)	Bone	0.93	0.920	0.93
	Chest	0.93	0.968	0.91
	MRI	0.93	0.904	0.95
VGG-16 8 blocks final layer quantized (99%)	Bone	0.946	0.931	0.959
	Chest	0.946	0.968	0.93
	MRI	0.946	0.940	0.95

For models where only the final layer is quantized using the maximum, KL divergence, and 99% methods, the results indicate relatively robust performance with minor variations across the different types of quantization. Accuracy remains consistently high, around 0.94 to 0.946, under the maximum and 99% quantization types, with slight decreases in the KL quantized version to 0.93. This minimal variance suggests that moderate quantization of the final layer does not drastically affect the model's ability to classify medical images accurately.

Precision and recall metrics show more variation. In the maximum quantization scenario, precision is high across all categories but slightly lower for MRI at 0.923, while recall is highest for MRI at 0.96, indicating good sensitivity. Under KL quantization, precision peaks for Chest images at 0.968, whereas recall is slightly reduced for Chest and MRI. The 99% quantization appears to balance both precision and recall effectively, particularly showing a notable improvement in Bone recall to 0.959.

However, a stark contrast is observed when the model is fully quantized. The accuracy decreases to 0.326 in all categories, while precision and recall also see a substantial decline, except for MRI recall which surprisingly stays high at 0.96. A significant decrease in performance highlights the challenges and possible drawbacks of full model quantization, especially in complex applications like as medical image classification, where accuracy and responsiveness to features are important.

The analysis underscores the necessity of careful consideration in the extent of quantization applied within neural networks, especially in medical settings where accuracy and reliability are crucial. Moderate quantization of the final layer preserves functionality, while full quantization severely impairs performance, suggesting that a balanced approach is essential for maintaining the efficacy of encrypted image retrieval systems.

VI. Discussion

In this discussion, we delve into the intricacies of implementing the VGG-16 model for encrypted and quantized image retrieval, focusing on the trade-offs between security, efficiency, and effectiveness. We explore the optimization of encryption and quantization methods to achieve optimal performance in a Content-Based Image Retrieval (CBIR) system.

1) The VGG-16 model was selected for the image retrieval task due to its deep architecture, which is highly effective in capturing complex textural details and spatial hierarchies in images. This characteristic is especially beneficial in medical imaging, where precise feature extraction is crucial for accurate classification and diagnosis. VGG-16's use of small, 3x3 convolutional kernels allows for a deeper network that can learn a rich hierarchy of features at multiple scales, making it a robust choice for both plaintext and encrypted image retrieval systems. The model's well-documented success in various image recognition tasks supports its adoption in exploring the impacts of encryption and quantization on medical image analysis.

2) The Role of Encryption and the Effectiveness and Security Balance Encryption plays a critical role in ensuring the confidentiality and security of sensitive data, such as medical images, in a CBIR system. However, as demonstrated by the performance metrics across different encryption granularities, there is a significant trade-off between security and the effectiveness of image retrieval. While high encryption levels (e.g., 1-block encryption) offer stronger data protection, they severely degrade the quality of features available for the model to learn and identify, leading to poorer classification accuracy and recall rates. Conversely, lighter encryption (e.g., 8-blocks) provides less security but maintains higher usability of the data, allowing for more effective feature extraction and image retrieval. Balancing these aspects requires a strategic approach to encryption that considers the specific needs for security versus the operational requirements for image analysis and retrieval.

3) Optimization of Encryption for Achieving Required Performance To optimize encryption for performance, it is essential to tailor the encryption methods to the specific characteristics of the data and the requirements of the retrieval system. For medical images, where detail and accuracy are paramount, adopting an encryption approach that preserves more structural and textural integrity—such as region-specific encryption or varying encryption levels based on the sensitivity of the image content—might be more suitable. Techniques such as homomorphic encryption could also be explored, as they allow certain types of computations to be carried out on encrypted data, thus maintaining privacy without sacrificing the ability to perform effective image retrieval. Additionally, optimizing the trade-off between block size in encryption and feature extraction needs is crucial, as smaller blocks increase security but reduce the quality of features, impacting model performance.

4) The Role of Quantization and the Efficiency and Effectiveness Trade-Off Quantization addresses the need to reduce the computational complexity and storage requirements of deep learning models, making them more suitable for deployment in resource-constrained environments such as mobile devices or in cloud-based architectures where computational efficiency is critical. The quantization of VGG-16, particularly at the final layer, has demonstrated the possibility of maintaining relatively high accuracy and precision with minor trade-offs in performance. However, fully quantized models exhibit a significant drop in performance, highlighting the delicate balance between efficiency and effectiveness. Quantization introduces noise and approximation errors into the model, which can degrade the accuracy if not managed carefully. The choice of quantization technique—maximum, KL divergence, or percentile-based—impacts how much of the critical information is preserved and thus the overall effectiveness of the model. Striking a balance between reducing computational demands and maintaining high classification accuracy is essential, particularly when dealing with high-stakes applications like medical diagnostics.

VII. Conclusion and Future Work

The fundamental problem of this study is in the development of a Content-Based Image Retrieval system that can effectively optimize the trade-off between privacy for encrypted images and computational efficiency and retrieval effectiveness. This is of utmost significance, particularly in fields where privacy and accuracy are closely related, such as in medical imaging. The proposed system is designed to efficiently handle an encrypted database of images, prioritizing the extraction of features and retrieval of images.

The primary objective of this study has been to investigate the VGG-16 architecture, which has gained recognition as a very efficient deep convolutional network for extracting high-level features from few imagine samples. Two essential modifications for preparing the model for encrypted domains were making adjustments to the granularity and applying quantization techniques in the last layer of the model. The primary objective of these approaches is to investigate the feasible and controlled impacts of encryption and quantization on the performance of a model.

Research findings indicated that VGG-16, in general, maintained a high level of accuracy in classifying plaintext images, but its performance declined as the level of detail arose. Therefore, it is the encryption with a granularity of 1-block that significantly and severely affects the ability of this model in accurately classifying images into particular classifications such as bone or MRI. The impact of quantization was considerably less severe but it became quite significant when the entire model was completely quantized. Research findings indicate that although a certain level of encryption and quantization can possibly be achieved with no impact on performance, over use of these techniques results in significant performance impairment.

One of the primary limitations of this study is the focus on a single model architecture (VGG-16) and a limited set of encryption and quantization strategies. Future research could explore other architectures like ResNet or EfficientNet, which might interact differently with encryption and quantization. Additionally, experimenting with newer encryption techniques like homomorphic encryption could offer insights into less disruptive methods. Further exploration into adaptive encryption and quantization strategies that dynamically adjust based on the content sensitivity and retrieval needs could also enhance system performance and usability.

Moreover, incorporating more sophisticated metrics for assessing the quality of retrieved images, such as structural similarity indexes or perceptual quality measures, could provide a deeper understanding of how encryption and quantization affect the perceived quality of images. Future studies should also consider the implementation of these systems in real-world applications, evaluating their practicality and efficiency in operational environments.

Acknowledgment

Communication of this research is made possible through monetary assistance by Universiti Tun Hussein Onn Malaysia and the UTHM Publisher's Office via Publication Fund E15216.

References

[1] S. Sikandar, R. Mahum, and A. Alsalman, A novel hybrid approach for a content-based image retrieval using feature fusion, vol. 13, no. 7, p. 4581, https://doi.org/10.3390/app13074581.

[2] G. Sumbul, J. Kang, and B. Demir, Deep learning for image search and retrieval in large remote sensing archives, pp. 150–160, https://doi.org/10.1002/9781119646181.ch11.

[3] M. S. Sayed, A. A. A. Gad-Elrab, K. A. Fathy, and K. R. Raslan, Unsupervised Content Based Image Retrieval Using Pre-Trained CNN and PCNN Features Extractors., vol. 16, no. 1, doi: 10.22266/ijies2023.0228.50.

[4] M. A. Mohammed, M. A. Hussain, Z. A. Oraibi, Z. A. Abduljabbar, and V. O. Nyangaresi, Secure Content Based Image Retrieval System Using Deep Learning, https://doi.org/10.56714/bj-rs.49.2.9.

[5] Z. Xia, L. Lu, T. Qiu, H. J. Shim, X. Chen, and B. Jeon, A Privacy-Preserving Image Retrieval Based on AC-Coefficients and Color Histograms in Cloud Environment., vol. 58, no. 1, doi:10.32604 /cmc.2019.02688.

[6] S S. H. Wady and R. Z. Yousif, A Secure Medical Image Transmission System Based On 2D Logistic Map and Diffie-Hellman Key Exchange Mechanisms, vol. 6, no. 2, pp. 94–104, https://doi.org/10.21928/uhdjst.v6n2y2022.pp94-104.

[7] S. R. Nair, High and Low-Level Classification based Features for Content-based Image Retrieval in Privacy-Preserving, https://doi.org/10.212-03/rs.3.rs-1461397/v1.

[8] F. Zhou, S. Qin, R. Hou, and Z. Zhang, Privacy-preserving image retrieval in a distributed environment, vol. 37, no. 10, pp. 7478–7501, https://doi.org/10.1002/int.22890.

[9] C. Zhang, C. Xu, K. Sharif, and L. Zhu, Privacy-preserving contact tracing in 5G-integrated and blockchain-based medical applications, vol. 77, p. 103520, https://doi.org/10.1016/j.csi.2021.1035-20.

[10] T. Janani and M. Brindha, Privacy-preserving transfer learning-based secure quantum image retrieval in encrypted domain for cloud environment, vol. 32, no. 2, p. 23003, https://doi.org/10.1117/1.JEI.32.2.023003.

[11] M. Naveen, A Review on Content Based Image Retrieval System Features Derived by Deep Learning Models, http://dx.doi.org/10.22214/ijraset.2021.39172.

[12] W. Wu, D. Li, J. Du, X. Gao, W. Gu, F. Zhao, X. Feng and H. Yan, An Intelligent Diagnosis Method of Brain MRI Tumor Segmentation Using Deep Convolutional Neural Network and SVM Algorithm, vol. 2020, p. 6789306, https://doi.org/10.1155/2020/6789306.

[13] Y. Zhang, Y. Ma, and Y. Liu, Convolution-Bidirectional Temporal Convolutional Network for Protein Secondary Structure Prediction, vol. 10, pp. 117469–117476, doi: 10.1109/ACCESS.2022.3219490.

[14] Y. Zhao and C. Feng, Remote Sensing Image Water Body Recognition Algorithm Based on Deep Convolution Generating Network and Combined Features, vol. 2022, p. 9932251, doi: 10.1155/2022/9932251.

[15] E. M. Alsaedi and A. kadhim Farhan, Retrieving Encrypted Images Using Convolution Neural Network and Fully Homomorphic Encryption, doi: 10.21123/bsj.2022.6550.

[16] Y. Wang, F. Wang, F. Liu, and X. Wang, Securing content-based image retrieval on the cloud using generative models, vol. 81, no. 22, pp. 31219–31243, https://doi.org/10.1007/s11042-022-12880-6.

[17] P. Desai, J. Pujari, C. Sujatha, A. Kamble, and A. Kambli, Hybrid approach for content-based image retrieval using VGG16 layered architecture and SVM: an application of deep learning, vol. 2, no. 3, p. 170, https://doi.org/10.1007/s42979-021-00529-4.

[18] C. I. Kerley, Y. Huo, S. Chaganti, S. Bao, M. B. Patel, and B. A. Landman, Montage Based 3D Medical Image Retrieval From Traumatic Brain Injury Cohort Using Deep Convolutional Neural Network, doi: 10.1117/12.2512559.

[19] A. Ahmed, A. O. Almagrabi, and A. H. Osman, Pre-Trained Convolution Neural Networks Models for Content-Based Medical Image Retrieval, doi: 10.21833/ijaas.2022.12.002.

[20] D. B. Mahesh, B. Madhuri, and R. L. D, Integration of Optimized Local Directional Weber Pattern With Faster Region Convolutional Neural

Network for Enhanced Medical Image Retrieval and Classification, doi: 10.1111/coin.12506.

[21] C. Palai, P. K. Jena, S. R. Pattanaik, T. Panigrahi, and T. Mishra, Content-Based Image Retrieval Using Encoder Based RGB and Texture Feature Fusion, doi: 10.14569/ijacsa.2023.0140328.

[22] C. G. Sotomayor, M. Mendoza, V. Castañeda, H. Farías, G. T. Molina, G. Pereira, S. Härtel, M. Solar and M. Araya, Content-Based Medical Image Retrieval and Intelligent Interactive Visual Browser for Medical Education, Research and Care, doi: 10.3390/diagnostics11081470.

[23] S. Bilquees, H. Dawood, H. Dawood, N. Majeed, A. Javed, and M. T. Mahmood, Noise Resilient Local Gradient Orientation for Content-Based Image Retrieval, vol. 2021, p. 4151482, doi: 10.1155/2021/4151482.

[24] X. Zhang and H. Cheng, Histogram-based retrieval for encrypted JPEG images, pp. 446–449, doi: 10.1109/ChinaSIP.2014.6889282.

[25] H. Cheng, X. Zhang, and J. Yu, AC-coefficient histogram-based retrieval for encrypted JPEG images, vol. 75, no. 21, pp. 13791–13803, doi: 10.1007/s11042-015-2741-z

[26] P. Li and Z. Situ, Encrypted JPEG image retrieval using histograms of transformed coefficients, no. November, pp. 1140–1144, doi: 10.1109/APSIPAASC47483.2019.9023179.

[27] H. Liang, X. Zhang, and H. Cheng, Huffman-code based retrieval for encrypted JPEG images, vol. 61, pp. 149–156, doi: 10.1016/j.jvcir.2019.03.021.

[28] Z. Xia, L. Wang, J. Tang, N. N. Xiong, and J. Weng, A Privacy-Preserving Image Retrieval Scheme Using Secure Local Binary Pattern in Cloud Computing, vol. 8, no. 1, pp. 318–330, doi: 10.1109/TNSE.2020.3038218.

[29] Z. Xia, L. Jiang, D. Liu, L. Lu, and B. Jeon, BOEW: A Content-Based Image Retrieval Scheme Using Bag-of-Encrypted-Words in Cloud Computing, vol. 15, no. 1, pp. 202–214, doi: 10.1109/TSC.2019.2927215

[30] C.-Y. Hsu, C.-S. Lu, and S.-C. Pei, Image feature extraction in encrypted domain with privacy-preserving SIFT, vol. 21, no. 11, pp. 4593–4607, doi: 10.1109/TIP.2012.2204272.

[31] L. Zhang, T. Jung, P. Feng, K. Liu, X. Y. Li, and Y. Liu, PIC: Enable large-scale privacy preserving content-based image search on cloud, vol. 2015-Decem, pp. 949–958, doi: 10.1109/ICPP.2015.104.

[32] R. Bellafqira, G. Coatrieux, D. Bouslimi, and G. Quellec, Content-based image retrieval in homomorphic encryption domain, in 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2015, pp. 2944–2947, doi: 10.1109/EMBC.2015.7319009.

[33] C. Guo, J. Jia, K. K. R. Choo, and Y. Jie, Privacy-preserving image search (PPIS): Secure classification and searching using convolutional neural network over large-scale encrypted medical images, vol. 99, p. 102021, doi: 10.1016/j.cose.2020.102021.

[34] B. Cheng, L. Zhuo, Y. Bai, Y. Peng, and J. Zhang, Secure index construction for privacy-preserving large-scale image retrieval, no. 4142009, pp. 116–120, doi: 10.1109/BDCloud.2014.36.

[35] B. Ferreira, J. Rodrigues, J. Leitão, and H. Domingos, Practical Privacy-Preserving Content-Based Retrieval in Cloud Image Repositories, vol. 7, no. 3, pp. 784–798, doi: 10.1109/TCC.2017.2669999.

[36] J. Huang, Y. Luo, M. Xu, S. Fu, and K. Huang, Accelerating Privacy-Preserving Image Retrieval with Multi-Index Hashing, pp. 492–497, doi: 10.1109/SEC54971.2022.00075.

[37] Y. Li, J. Ma, Y. Miao, Y. Wang, X. Liu, and K. K. R. Choo, Similarity Search for Encrypted Images in Secure Cloud Computing, vol. 10, no. 2, pp. 1142–1155, doi: 10.1109/TCC.2020.2989923.

[38] Y. Li, J. Ma, Y. Miao, H. Li, Q. Yan, Y. Wang, X. Liu and K. K. R. Choo, DVREI: Dynamic Verifiable Retrieval Over Encrypted Images, vol. 71, no. 8, pp. 1755–1769, doi: 10.1109/TC.2021.3106482.

[39] D. Liu, J. Shen, Z. Xia, and X. Sun, A content-based image retrieval scheme using an encrypted difference histogram in cloud computing, vol. 8, no. 3, p. 96, https://doi.org/10.3390/info8030096.

[40] W. Lu, A. L. Varna, and M. Wu, Confidentiality-preserving image search: A comparative study between homomorphic encryption and distance-preserving randomization, vol. 2, pp. 125–141, doi: 10.1109/ACCESS.2014.2307057.

[41] W. Lu, A. Swaminathan, A. L. Varna, and M. Wu, Enabling search over encrypted multimedia databases, in Media Forensics and Security, International Society for Optics and Photonics, 2009, p. 725418, doi : 10.1117/12.806980.

[42] Q. Zou, J. Wang, and X. Chen, Secure Encrypted Image Search in Mobile Cloud Computing, pp. 572–575, https://doi.org/10.1109/BWCCA.201-5.41.

[43] H. Cheng, X. Zhang, J. Yu, and F. Li, Markov process-based retrieval for encrypted JPEG images, vol. 2016, no. 1, pp. 1–9, doi: 10.1186/s13635-015-0028-6.

[44] D. Liu, J. Shen, Z. Xia, and X. Sun, A content-based image retrieval scheme using an encrypted difference histogram in cloud computing, vol. 8, no. 3, p. 96, https://doi.org/10.3390/info8030096.

[45] P. Yu, J. Tang, Z. Xia, Z. Li, and J. Weng, A Privacy-Preserving JPEG Image Retrieval Scheme Using the Local Markov Feature and Bag-of-Words Model in Cloud Computing, vol. 11, no. 3, pp. 2885–2896, doi: 10.1109/TCC.2022.3233421.

[46] Y. Ma, X. Chai, Z. Gan, and Y. Zhang, Privacy-preserving TPE-based JPEG image retrieval in cloud-assisted internet of things, doi: 10.1109/JIOT.2023.3301042.

[47] K. Jayashree, Secureimagesec: A privacy-preserving framework for outsourced picture representation with content-based image retrieval, no. Preprint, pp. 1–22, doi: 10.3233/IDA-240265.

[48] S. Kumar, A. K. Pal, S. K. H. Islam, and M. Hammoudeh, Secure and efficient image retrieval through invariant features selection in insecure cloud environments, vol. 35, no. 7, pp. 4855–4880, https://doi.org/10.1007/s00521-021-06054-y.

[49] G. Sucharitha, D. Godavarthi, J. V. N. Ramesh, and M. I. Khan, Secure and efficient content-based image retrieval using dominant local patterns and watermark encryption in cloud computing, pp. 1–17, https://doi.org/10.1007/s10586-024-04635-9.

[50] P. Li, L. Han, X. Tao, X. Zhang, C. Grecos, A. Plaza and P. Ren, Hashing Nets for Hashing: A Quantized Deep Learning to Hash Framework for Remote Sensing Image Retrieval, vol. 58, no. 10, pp. 7331–7345, doi: 10.1109/TGRS.2020.2981997.

[51] B. Yang, L. Yang, X. Li, W. Zhang, H. Zhou, Y. Zhang, Y. Ren and Y. Shi, 2-bit Model Compression of Deep Convolutional Neural Network on ASIC Engine for Image Retrieval, https://doi.org/10.48550/arXiv.1905.03362.

[52] X. Zhang, C. Bai, and K. Kpalma, OMCBIR: Offline mobile content-based image retrieval with lightweight CNN optimization, vol. 76, p. 102355, doi: 10.1016/j.displa.2022.102355.

[53] J. Chu, L. Li, and X. Xiao, Remote Sensing Image Retrieval by Multi-Scale Attention-Based CNN and Product Quantization, vol. 2021-July, pp. 8292–8297, doi: 10.23919/CCC52363.2021.9549732.

[54] Z. Wei, K. Jin, Z. Zhang, and X. Zhou, Multi-label contrastive hashing, vol. 149, no. August 2023, p. 110239, doi: 10.1016/j.patcog.2023.110239.

[55] Y. Liu, Y.-H. Zhu, X. Song, J. Song, and D.-J. Yu, Why can deep convolutional neural networks improve protein fold recognition? A visual explanation by interpretation, vol. 22, no. 5, p. bbab001, doi: 10.1093/bib/bbab001.

[56] D. Zhu, Y. Fu, X. Zhao, X. Wang, and H. Yi, Facial Emotion Recognition Using a Novel Fusion of Convolutional Neural Network and Local Binary Pattern in Crime Investigation, vol. 2022, p. 2249417, doi: 10.1155/2022/2249417.

[57] H. Afzaal, A. A. Farooque, A. W. Schumann, N. Hussain, A. McKenzie-Gopsill, T. Esau, F. Abbas and B. Acharya, Detection of a Potato Disease (Early Blight) Using Artificial Intelligence, doi: 10.3390/rs13030411.

Tracking Computer Vision Algorithm Based on Fusion Twin Network

Xin Wang
Software Institute, Hunan College of Information, Changsha, 410016, China

***Abstract*—Deep learning technology has promoted the rapid development of visual object tracking, among which algorithms based on twin networks are a hot research direction. Although this method has broad application prospects, its performance is often greatly reduced when encountering target occlusion or similar objects in the background. In response to this issue, a method is proposed to integrate channel and spatial dimension attention mechanisms into the backbone architecture of twin networks, to optimize the algorithm's recognition accuracy for tracking targets and its stability in changing environments. Then, a region recommendation network based on adaptive anchor box generation is adopted, combined with twin networks to enhance the network's modeling ability for complex situations. Finally, a new visual tracking algorithm is designed. Through comparative experiments, the success rate of the former increased by 0.6% and 0.9% respectively on the two datasets, and its accuracy also increased by 1.2% and 1.8% accordingly. The success rate of the latter increased by 1.5% and 1.2% respectively in the two datasets, and the accuracy also increased by 1.2% and 0.6% respectively. From this, the improved algorithm can improve the performance of target tracking and has certain application potential in visual target tracking.**

***Keywords*—Visual tracking; twin network; integration; attention mechanism; self-adaption**

I. Introduction

Fatigue driving has become an important factor in causing traffic accidents, posing a serious threat to social safety and public health. Currently, although there are various fatigue driving detection models based on computer algorithms, they still face many challenges in practical applications, such as poor comfort, susceptibility to external factors such as lighting, masks, sunglasses, low detection accuracy, and poor real-time performance [1]. With the successful application of deep learning in object detection, some studies have begun to use detection techniques to guide the development of object tracking technology [2]. Xin et al. proposed a new Siamese adaptive learning network for visual tracking to address the manually adjusting parameters. The designed method took spatial alignment and model learning state as criteria for anchor quality evaluation, and employed Gaussian mixture distribution for adaptive allocation instead of IoU-based anchor allocation. The experimental results showed that the tracker had superiority on benchmarks GOT-10k, and LaSOT [3]. In order to address the issue of retaining much unfavorable background information in association operations, Jun W et al. proposed an effective feature recognizer that included channel and spatial attention modules to focus on key information. The representation capability was optimized. Experiments on six benchmark tests showed that the designed tracker outperformed other trackers. Especially, it achieved 80.4% AUC on TrackingNet and 68.4% AUC on GOT-10k during real-time operation [4]. Due to the insufficient control accuracy and stability of current virtual vision tracking technology, Jianbin et al. designed a new intelligent algorithm on the basis of human-computer interaction and virtual vision tracking technology to improve the overall performance of virtual vision tracking. The designed model was efficient, with tracking accuracy more than 10% higher than traditional methods [5]. Di et al. found that the discriminative model for predicting object tracking models was susceptible to interference from similar objects and required much-labeled data for training during actual use. Therefore, two methods were proposed to enhance the robustness of target tracking against interference from similar objects: multi-scale region search and response map processing based on Gaussian convolution. A large number of experiments showed that the enhancement function implemented in the tracking framework enhanced its robustness. The tracker based on self-supervised training had excellent tracking performance [6]. The existing Siamese trackers are insufficient to effectively distinguish between targets and the fluctuation interference embedded in the two branches of information, resulting in inaccurate classification and localization. Therefore, Liu et al. proposed two novel sub-network spaces for spatial feature embedding to optimize the discriminative ability of trackers in the embedding space and their adaptability to complex tracking scenarios. Compared with the most advanced trackers, the proposed tracker had competitive tracking performance in background clutter and similar object attributes, verifying the effectiveness of the method [7]. To solve the tracking drift caused by inaccurate initial positions in most existing trackers, Han et al. proposed a deep learning method that could generate accurate positions of objects given their rough positions. The proposed method was applied in the tracking process to improve the accuracy. A large number of experiments in object tracking benchmark testing verified its effectiveness [8]. Due to the lack of attention on the channel dimension in current trackers, their potential tracking capabilities are hindered. Therefore, Shaochuan et al. used a novel spatial channel converter that integrated information conveyed by features along both spatial and channel directions. To quantify temporal smoothness, a jitter metric that measured the cross-frame variation of predicted bounding boxes was proposed as a function of parameters like center displacement, area, and aspect ratio. Several well-known benchmark datasets demonstrated its robustness [9].

In the field of computer vision tracking, although some progress has been made in research, fatigue driving detection

technology still faces multiple challenges. Firstly, existing models have shortcomings in terms of comfort, which may cause unnecessary interference to drivers. Secondly, sensitivity to external factors such as lighting, masks, sunglasses, etc. affects the accuracy and robustness of detection. In addition, detection accuracy and real-time performance are also key areas that need to be improved in current technology. For example, although the Siamese adaptive learning network proposed by Xin et al. performed well on GOT-10k and LaSOT benchmark tests, further optimization is still needed to improve tracking accuracy and stability. The effective feature recognizer proposed by the military and others has optimized its representation ability, but there is still room for improvement in real-time performance. The new spatial feature embedding subnetwork proposed by Liu et al. has optimized the recognition ability of trackers, but its adaptability in complex tracking scenarios still needs further verification. Although the deep learning method proposed by Han et al. can generate accurate positions of objects, there is still room for improvement in dealing with tracking drift caused by inaccurate initial positions.

To address these drawbacks, this study proposes a visual tracking algorithm based on traditional dual networks, combining attention mechanism and adaptive anchor box generation to improve the algorithm performance. The attention mechanism helps the algorithm to focus on key information and improve the efficiency of target feature extraction, while the adaptive anchor box generation is able to evaluate the anchor quality more accurately, using a Gaussian mixture distribution for adaptive allocation, thus improving the target tracking performance in complex environments.

II. Computer Vision Algorithm Based on Twin Network Tracking

A. Twin Network Visual Tracking Based on Attention Mechanism

In fatigue driving detection, the characteristic of computer vision algorithms is that they can analyze the driver's behavior and physiological signals in real time, such as blink frequency, head posture, etc., to determine whether they are in a fatigue state. However, existing algorithms often have low detection accuracy due to factors such as background interference and lighting changes [10]. To this end, an improved algorithm is proposed by introducing an attention mechanism to extract useful features. The designed structure utilizes multi-layer linear fusion to integrate different levels of feature information, thereby enhancing the recognition accuracy and robustness for fatigue states. Its structure is shown in Fig. 1.

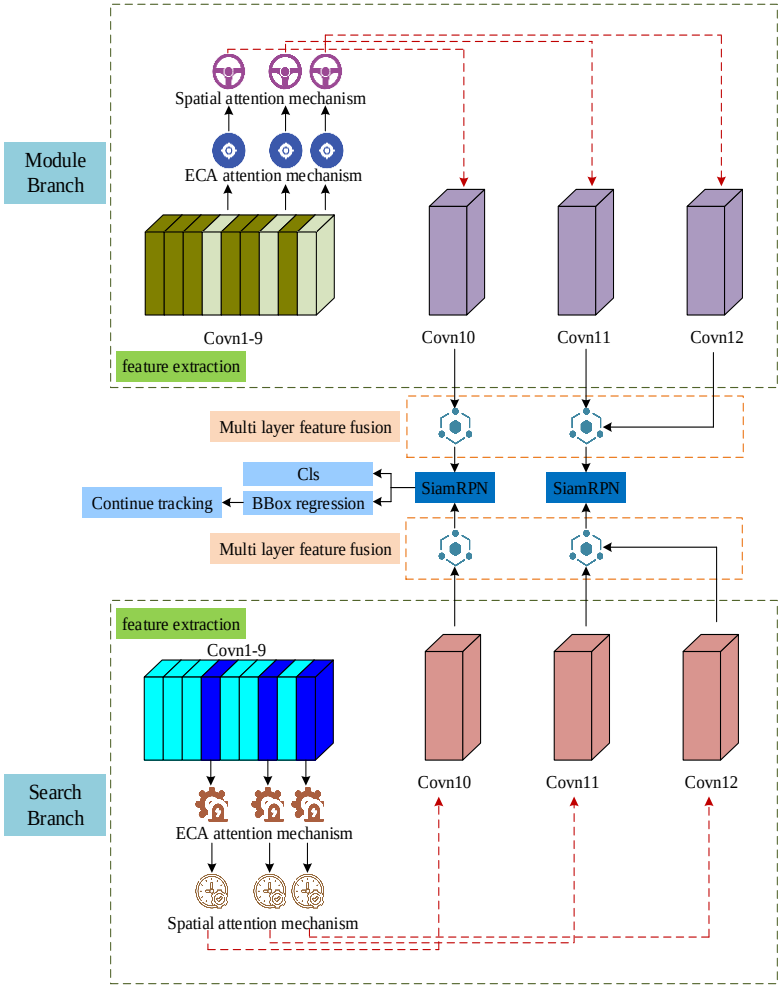


Fig. 1. Structure diagram of twin network target tracking algorithm incorporating attention mechanism.

From Fig. 1, the algorithm optimizes the feature extraction process in target tracking tasks by integrating channel and spatial attention mechanisms, thereby improving the performance of the tracking algorithm. Through analysis, it is found that striking a balance between accuracy and efficiency in target tracking is crucial [11]. Specifically, the research adopts the SiamRPN++ framework and makes improvements and adjustments based on it. Firstly, for the optimization of feature extraction, this method is similar to a fine filter. It can filter out feature information that is not conducive to target tracking, and only retain key information that helps improve tracking performance. This ensures the significant differences in feature responses across channels and spaces, providing a more accurate basis for subsequent similarity calculations. Secondly, in order to improve algorithm performance, effective fusion of feature information can be achieved through improvements to the backbone network. It includes two main parts: module branch and search branch. Both branches use Covin1-9 as the initial feature extraction layer and apply ECA attention mechanism and spatial attention mechanism to optimize feature representation capability. In the module branch, the features after passing through Covin1-9 are sent to the SiamRPN module for target tracking, and then the BBox regression is used to further accurately locate the position of the target [12]. In addition, the module branch also performs multi-layer feature fusion operations, combining features of different depths to obtain richer information. In the search branch, multi-layer feature fusion operations are also performed, but the fusion is performed after Covn1-9, indicating that this branch may be more focused on extracting and processing low-level features. This design fully utilizes the advantages of attention mechanism, enabling the network to output more accurate feature information, as shown in Fig. 2.

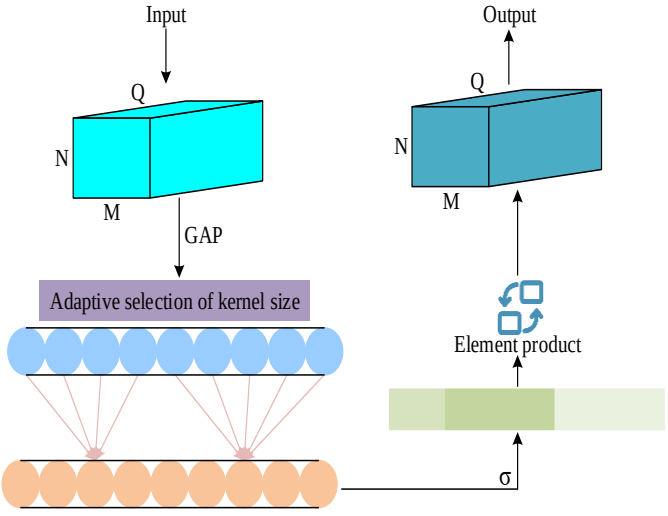


Fig. 2. Structure diagram of channel attention mechanism module.

From Fig. 2, the main purpose of the channel attention mechanism is to capture key feature information. It significantly enhances the network's performance in feature extraction by assigning higher weights to the target feature channel [13]. Then, the weights of each channel are obtained through the activation function, as shown in Eq. (1).

$$f(z)=\frac{1}{1+e^{-z}} \tag{1}$$

In Eq. (1), $f(z)$ represents the activation function. z represents the initial weight. The channel attention related feature map is shown in Eq. (2).

$$F_{ECA,i}=\omega_i \times F_i \tag{2}$$

In Eq. (2), $F_{ECA,i}$ is the attention feature map. ω_i is the weight on channel i. F_i is the original feature map on channel i. The output ω_i of each channel is weighted and summed all the features y_i^j within its receptive field, and then generated by the activation function σ, as shown in Eq. (3).

$$\omega_i=\sigma\left(\sum_{j=1}^{k}\alpha_i^j y_i^j\right), y_i^j \in \Omega_i^k \tag{3}$$

In Eq. (3), y_i^j represents the j-th element in Ω_i^k. α_i^j signifies the weight of the j-th element in the i-th channel. These weights are obtained through learning and are used for weighted combinations of different features. Then the corresponding weight of the size is generated, as shown in Eq. (4).

$$\omega=\sigma\left(CID_k(y)\right) \tag{4}$$

In Eq. (4), CID is a one-dimensional convolution. The size of k determines the coverage of interaction, and its relationship with the number of channels is shown in Eq. (5).

$$D \approx \exp(\gamma \times k - b) \tag{5}$$

In Eq. (5), D signifies the channels. γ and b represent empirical values. k can be determined by the number of channels D, as shown in Eq. (6).

$$k=\left|\frac{\log_2(D)}{\gamma}+\frac{b}{\gamma}\right| \tag{6}$$

To reduce the impact of background noise on target tracking and make the network more focused on the target itself. The study introduces the spatial attention mechanism, which weights different spatial positions of the feature map to focus on the key areas of the target object, while suppressing interference from background or irrelevant areas [14-15]. This helps the network automatically learn and focus on the key spatial positions of the target object, such as edges, corners, etc., thereby improving the recognition and tracking accuracy of the target. Its structure is shown in Fig. 3.

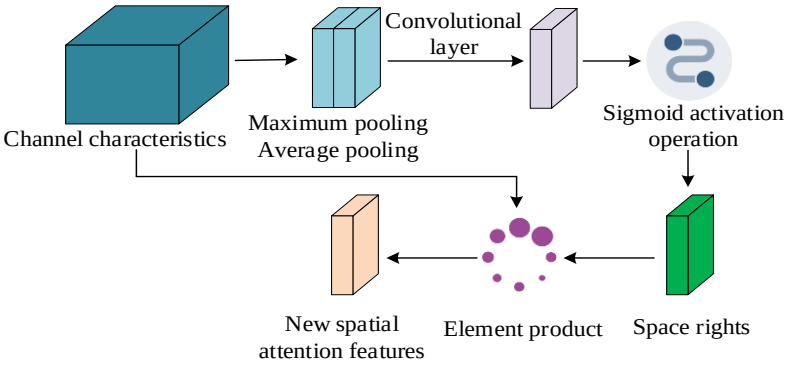


Fig. 3. Structure diagram of spatial attention mechanism module.

From Fig. 3, the input feature map first extracts deep features through convolutional layers, and then normalizes the output using the Sigmoid activation function, laying the foundation for subsequent spatial attention operations. The spatial attention mechanism multiplies the Sigmoid activated output with the original feature map through element multiplication operation. This operation achieves spatial position weighting of the feature map, focusing on key areas of the target (such as edges and corners) while reducing the interference of background noise. The weighted feature map is further processed through convolutional layers and combined with max pooling and average pooling techniques to extract global feature information. After these features are fused, they are refined again through convolutional layers. The generated spatial feature map significantly enhances the representation ability of the target area.

Finally, the spatial attention features is shown in Eq. (7).

$$M_S(F)=\sigma\left(f^{7*7}\left(\left[F_{avg}^{s};F_{\max}^{s}\right]\right)\right) \tag{7}$$

In Eq. (7), σ is the activation operation. F_{avg}^{s} is the average pooling feature. $F_{\max}^{s}$ is the maximum pooling feature. Then, multi-layer feature linear fusion is performed, and the fusion process is shown in Eq. (8).

$$\begin{cases}\hat{F}_4=\sum\left(F_4,F_5\right)\\ \hat{F}_3=\sum\left(F_3,\hat{F}_4\right)\end{cases} \tag{8}$$

In Eq. (8), $\hat{F}_4$ represents the fused mid-level feature map. F_4 represents the mid-level feature map. F_5 represents the deep feature map. $\hat{F}_3$ represents the shallow feature map after fusion. F_3 represents the shallow feature map.

B. *Twin Network Visual Tracking Algorithm Integrating Adaptive Anchor Box Generation*

Although introducing attention mechanisms and multi-layer linear fusion can optimize the extraction accuracy and the robustness of algorithms, there are still some potential drawbacks. Firstly, the attention mechanism may increase the computational burden of the model, especially when dealing with rich channel and spatial position weights. Secondly, the model may overfit the training data, resulting in candidate boxes that cannot cover specific targets, thereby reducing its generalization ability on unseen data. Therefore, this study introduces a new adaptive anchor box to reduce manual intervention, enabling the model to automatically learn and generate candidate boxes based on the semantic information of the image itself, while reducing the computational burden of the attention mechanism. Its structure is displayed in Fig. 4.

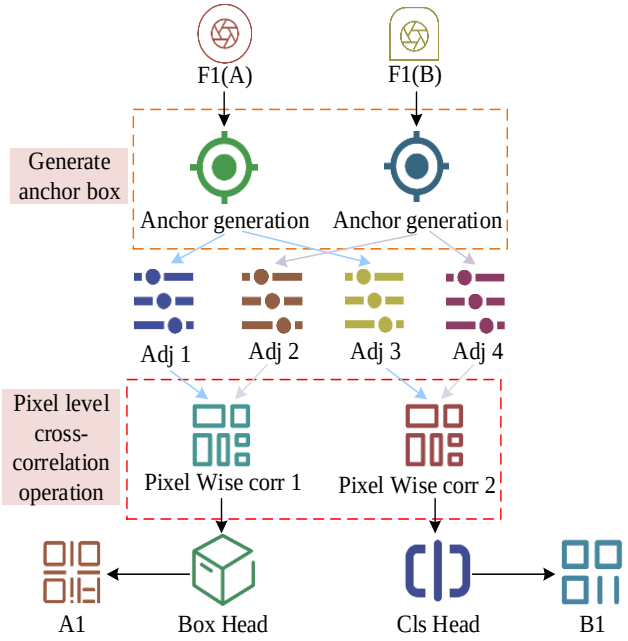


Fig. 4. Network structure diagram of adaptive generation anchor box region recommendation.

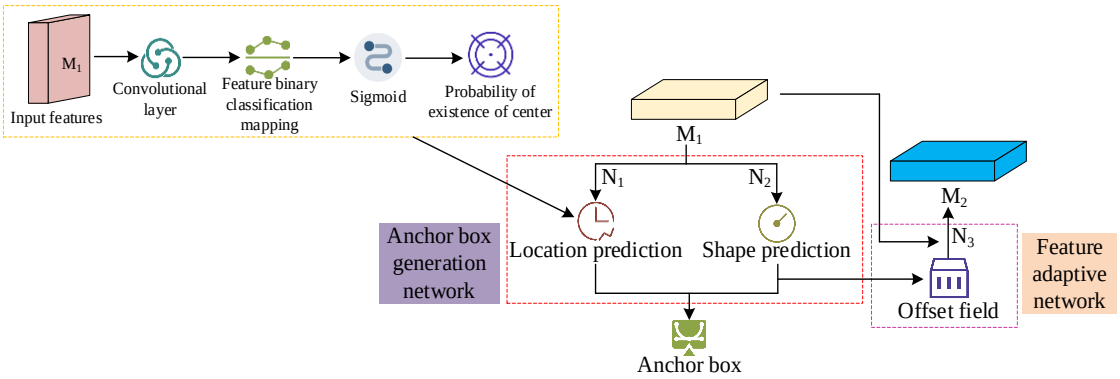


Fig. 5. Structure diagram of learnable generation module.

From Fig. 4, the input image generates feature maps through two convolutional layers F1 (A) and F1 (B), which are used for anchor box generation. Anchor boxes are generated at multiple scales (Adj 1 to Adj 4) through preset proportions and sizes to accommodate targets of different sizes and shapes. Subsequently, pixel level cross-correlation operations are performed between feature maps to generate Pixel Wise corr 1 and Pixel Wise corr 2 feature maps by calculating the correlation of corresponding pixel points, enhancing the spatial position information and correlation between features. These processed feature maps are further fed into the Box Head and Cls Head modules. Box Head is responsible for accurately regressing the coordinates of the target bounding box, while Cls Head performs classification tasks and predicts the target category for each anchor box. By optimizing the position and size of the anchor box, a detection box is ultimately generated to complete object detection. Based on the finer grained cross-correlation operations, the generated feature boundaries are made clearer [16]. This improves the success rate and accuracy of target tracking, thereby achieving more precise target tracking results. The core structure of generating anchor boxes is shown in Fig. 5.

From Fig. 5, the module has two parts: one is the anchor box generation network, which is used to generate unique anchor boxes, and the other is the feature adaptive network. The input features are convolved to obtain a binary classification map, and then the probability of center existence is obtained through the Sigmoid activation function; Next, this probability value is input into the anchor box generation network to generate anchor boxes. At the same time, the input features are fed into the feature adaptive network to obtain three features: N1, N2, and N3. Finally, these three features are input together with the anchor box into the localization prediction and shape prediction to obtain the final offset field. This method can infer the position and shape of the anchor box, as displayed in Eq. (9).

$$p(x,y,w,h \mid I) = p(x,y \mid I)\, p(w,h \mid x,y,I) \tag{9}$$

In Eq. (9), $p(x,y,w,h \mid I)$ represents the position and shape of the target. $p(x,y \mid I)$ represents that anchor boxes have different probabilities of appearing at different positions [17]. Convolutional neural networks automatically learn feature representations of images through a series of convolutional layers. The study employs a two channel convolutional sub network, and these features are subsequently used for classification tasks. Two channels independently predict the width and height of the corresponding anchor box, allowing the predicted box to more accurately fit the actual size of the target object, as displayed in Eq. (10).

$$\begin{cases} w = \sigma * s * e^{dw} \\ h = \sigma * s * e^{dh} \end{cases} \tag{10}$$

In Eq. (10), w is the width. h is high. σ is the empirical value. s is the step size. This study conducts deformation convolution with offset to obtain adaptive feature maps based on anchor box shapes at different positions, as shown in Eq. (11).

$$f_i^{'} = N_3\left(f_i, w_i, h_i\right) \tag{11}$$

In Eq. (11), $f_i^{'}$ is the convolved feature. N_3 is the deformable convolution. f_i is the original feature. (w_i, h_i) is the shape of the anchor box corresponding to the position. In the application of adaptive anchor box region recommendation network, the loss function is one of the key factors in the training process, which is applied to measure the difference between the predicted and the true situation [18-19]. Therefore, the study combines learning position and shape branches to form an adaptive anchor box loss function. By minimizing this comprehensive loss, the model can learn more accurate object detection and tracking capabilities [20]. Its expression is shown in Eq. (12).

$$L = \lambda_1 L_{loc} + \lambda_2 L_{shape} + L_{cls} + L_{reg} \tag{12}$$

In Eq. (12), L_{loc} is the position loss, which is used to measure the difference in position (i.e. the coordinates of the bounding box) between the predicted box and the real box. L_{shape} signifies the shape loss, used to measure the difference in shape (i.e. the aspect ratio of the bounding box) between the predicted box and the real box. By optimizing this loss, the model can better learn the shape features of the target object. L_{cls} is the classification loss, and L_{reg} is the regression loss, used in object tracking to measure the distance between the

predicted and the true boxes. λ_1 and λ_2 are used to balance the relative importance of position loss and shape loss in the total loss. In the position prediction branch, the loss function is displayed in Eq. (13).

$$L_{loc} = -\alpha(1-p)^{\gamma}\log(p) \tag{13}$$

In Eq. (13), α and γ represent the empirical values of the parameters. p signifies the probability of positive samples. In the shape prediction branch, the loss function can be used to optimize shape prediction without calculating the objective, as shown in Eq. (14).

$$L_{shape} = L_1\left(1-\min\left(\frac{w}{w_g},\frac{w_g}{w}\right)\right)+L_1\left(1-\min\left(\frac{h}{h_g},\frac{h_g}{h}\right)\right) \tag{14}$$

In Eq. (14), L_1 represents the smoothing loss. w_g signifies the real frame width, and h_g signifies the height. w signifies the candidate box width, and h signifies the height. After generating adaptive anchor boxes, the study will also use pixel level cross-correlation operations to represent features with higher quality, as shown in Fig. 6.

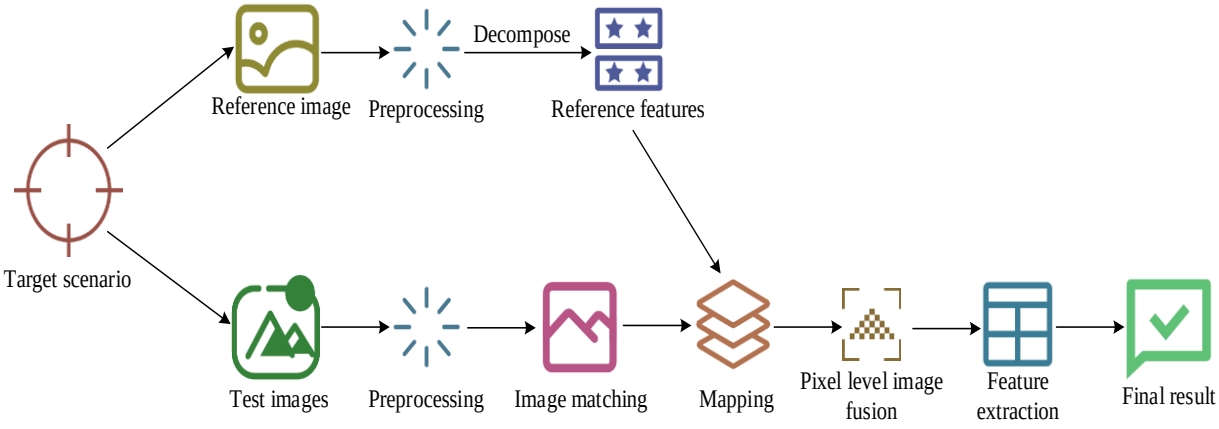


Fig. 6. Pixel level cross-correlation specific process.

From Fig. 6, in the object detection process, comparing the reference features with the test features generated after pre-processing of the test image, each part acts as a kernel to achieve deep interaction at the pixel level. After pixel level cross-correlation operation, high-quality features are sent to the mapping module for further processing. After optimization by the feature extraction module, detection results containing the precise position and category of the target are generated to avoid the target feature blurring. The process of generating fused images is shown in Eq. (15).

$$A = \left\{ A_i \mid A_i = Z_i * X \right\}_{i\in\{1,..,n\}} \tag{15}$$

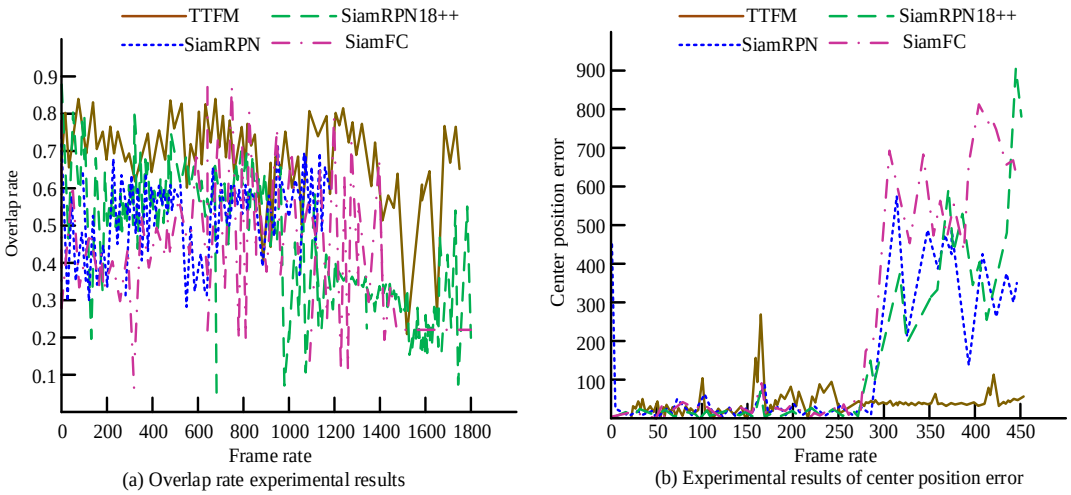
In Eq. (15), Z_i is the decomposed convolution kernel. $*$ signifies the convolution operation. X signifies the feature of the search area.

III. Results and Discussion

A. Analysis of Twin Network Visual Tracking Based on Attention Mechanism

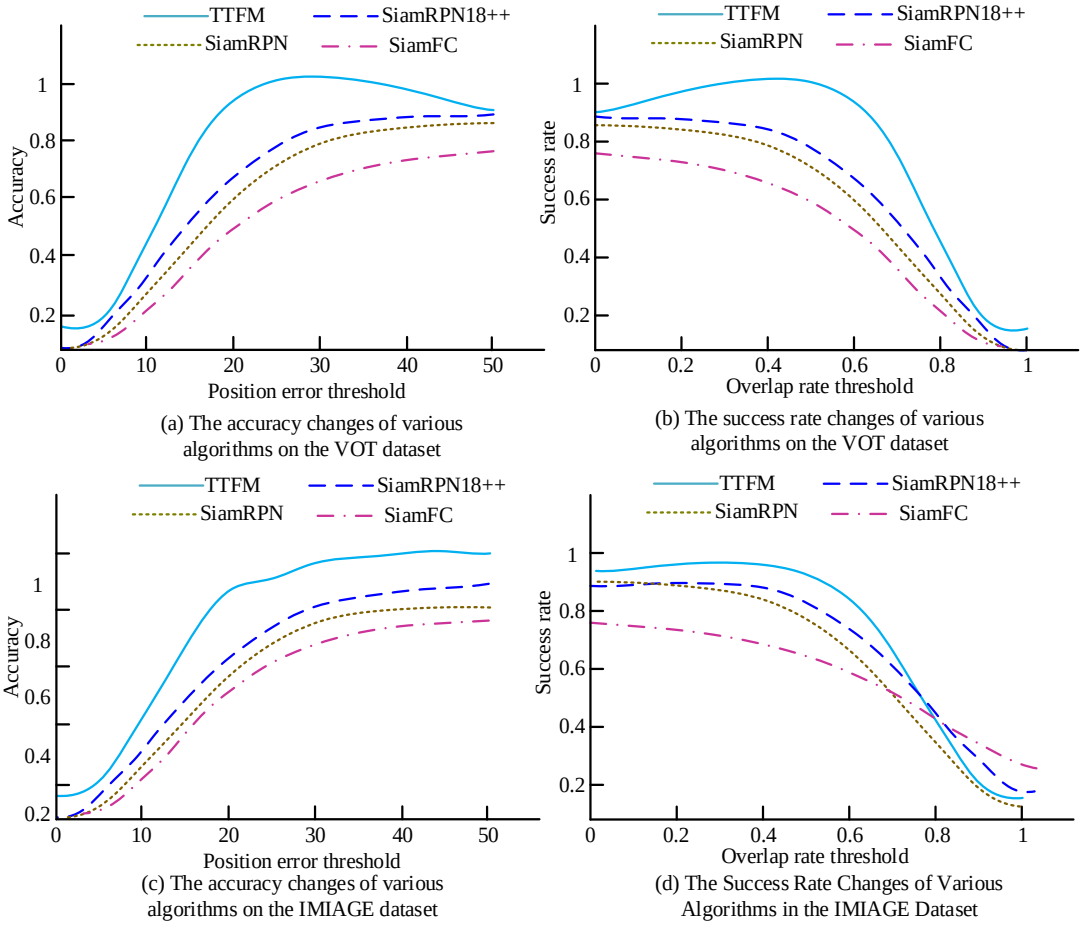
In order to solve the low accuracy and poor robustness in traditional fatigue driving recognition, a twin network visual tracking algorithm on the basis of attention mechanism is proposed. To verify the performance of the proposed TTFM in target tracking in complex environments, particularly its robustness against interference factors such as lighting changes, occlusion, and size variations, the study first conducts multiple comparative experiments to assess the advantages of TTFM algorithm in tracking accuracy and efficiency. Fig. 7 displays the results, taking the center position error of overlap rate as the evaluation index.

From Fig. 7, the TTFM algorithm had the best overlap rate of 0.75, indicating the best performance. Compared with other algorithms, the optimized TTFM algorithm increased the average overlap rate by 0.156%, effectively optimizing the efficiency of feature expression and target feature extraction. This lays the foundation for improving target tracking performance in complex environments. The TTFM algorithm performed well in tracking error at the center position of the sequence, with the smallest error. Compared with other basic algorithms, the TTFM algorithm reduced the center position error by 5.84, 30.63, and 43.42 pixels, respectively. When the target was subject to complex background interference such as changes in lighting, occlusion, and size, the TTFM algorithm significantly reduced the center position error. To further validate the effectiveness and performance improvement of the TTFM algorithm in single target tracking tasks. This study tested the commonly used VOT dataset and IMIAGE dataset, which contain visual tracking data under different adverse conditions such as lighting, to evaluate the improvement of TTFM algorithm and its generalization ability. The results are shown in Fig. 8.



(a) Overlap rate experimental results (b) Experimental results of center position error

Fig. 7. Comparison analysis curve of overlap rate and center position.



(a) The accuracy changes of various algorithms on the VOT dataset

(b) The success rate changes of various algorithms on the VOT dataset

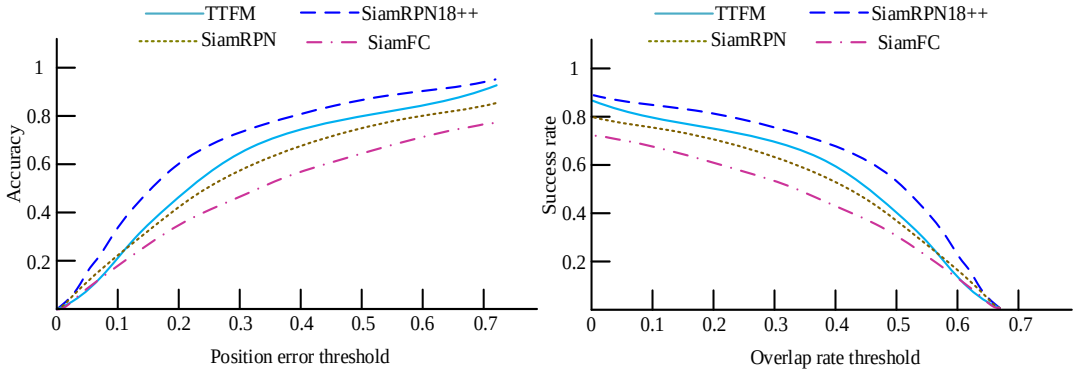
(c) The accuracy changes of various algorithms on the IMIAGE dataset

(d) The Success Rate Changes of Various Algorithms in the IMIAGE Dataset

Fig. 8. Success rate and accuracy on the VOT dataset.

From Fig. 8, the TTFM algorithm had excellent performance on the VOT dataset. Compared with the other three methods, the TTFM achieved obvious improvements in overall success rate and accuracy. Specifically, on the VOT dataset, the overall success rate of the TTFM algorithm reached 0.504, with an increase of 0.6% compared with the average success rate of other algorithms. Similarly, its accuracy also reached 0.642, with an increase of 0.9% compared with the average accuracy of other algorithms. However, the four algorithms do not change much in the IMIAGE dataset dataset,

indicating that these four algorithms have good generalization ability. The outstanding performance of the TTFM algorithm on the VOT dataset is mainly attributed to the feature extraction and fusion strategy proposed in the research. The algorithm significantly enhances the ability to extract useful features by introducing attention mechanisms, which is particularly important in complex scenarios. Secondly, the algorithm utilizes multi-layer linear fusion technology to integrate feature information from different levels, which not only enhances the richness of features, but also improves the accuracy and robustness in identifying fatigue states. To further validate the generalization ability of these algorithms, additional tests are conducted on the TrackingNet dataset, keeping hyper-parameters such as learning rate unchanged. Fig. 9 presents the experimental results.



(a) Accuracy (b) Success rate

Fig. 9. Success rate and accuracy on the TrackingNet dataset.

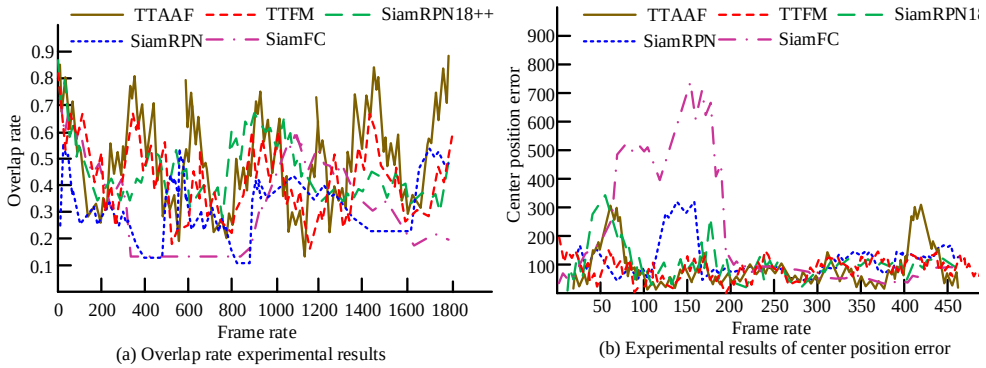
From Fig. 9, the TTFM performed better than the other three methods. For accuracy, the TTFM algorithm performed best when the position error threshold was low, ultimately converging to 0.830, while SiamFC performed better at higher position error thresholds. In the success rate, the TTFM algorithm performed well when the overlap rate threshold was high, while SiamFC performed well at lower overlap rate thresholds. The success rate and accuracy of the TTFM algorithm were 0.542 and 0.763, which were not significantly different from their performance in the VOT dataset. Overall, the TTFM algorithm performed the best on the TrackingNet dataset, especially for data containing occlusion attributes, averaging 0.4% and 3.3% higher than other algorithms. Compared with the VOT dataset, the performance of these three methods varied significantly. TTFM demonstrates superior tracking performance in complex environments. Therefore, the TTFM has better tracking performance and generalization ability, which is suitable for various complex tracking scenarios.

B. Analysis of Twin Network Visual Tracking Algorithm Integrating Adaptive Anchor Box Generation

From the analysis of the twin network visual tracking on the basis of attention mechanism, the tracking algorithm for preliminary components has good performance. Based on this, the study further optimizes it by introducing adaptive anchor box generation. To verify the repeatability and center position error of the twin network visual tracking algorithm that integrates adaptive anchor box generation, comparative analysis experiments are carried out on the VOT dataset, as presented in Fig. 10.

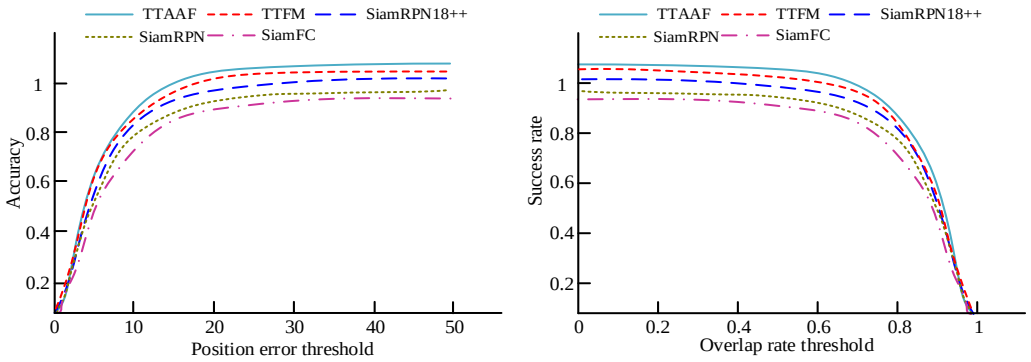
From Fig. 10, the improved algorithm had the best overlap rate. The improved algorithm performed well in terms of overlap rate, with an average overlap rate increase of 0.126 and an average center position error reduction of 23.450 pixels. The excellent performance of the improved algorithm in terms of overlap rate is mainly attributed to its innovative structure and mechanism. Firstly, the algorithm enhances the extraction of key behavioral and physiological signal features, such as blink frequency and head posture, by introducing attention mechanisms. These features are crucial for determining whether the driver is fatigued. Secondly, the application of multi-layer linear fusion effectively integrates feature information from different levels, enhancing the richness of features and the robustness of algorithms. To verify the tracking performance, comparative experiments are performed on the VOT dataset, as shown in Fig. 11.

From Fig. 11, the TTAAF algorithm performed best when the position error threshold was low, while SiamFC performed better when the position error threshold was high. In the success rate, the TTAAF algorithm performed well when the overlap rate threshold was high, while SiamFC performd well at lower overlap rate thresholds. Overall, the TTAAF algorithm performed the best on the VOT dataset, with an overall success rate and accuracy of 0.831 and 0.862, respectively, which were on average 0.3% and 0.5% higher than other algorithms. Especially for data containing occlusion attributes, the success rate and accuracy of the TTAAF algorithm were 0.642 and 0.753, which were on average 0.2% and 0.4% higher than other algorithms. Therefore, the TTAAF has better tracking performance and generalization ability, which is suitable for various complex tracking scenarios. In addition, the results of the four algorithms on the TrackingNet dataset are shown in Fig. 12.



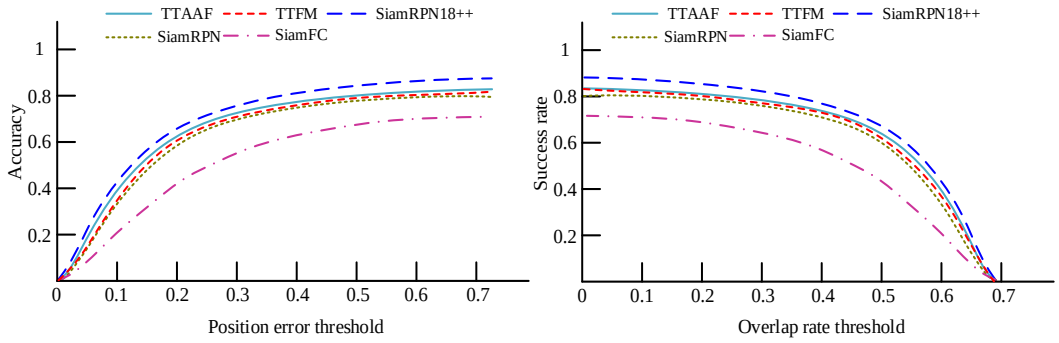
(a) Overlap rate experimental results (b) Experimental results of center position error

Fig. 10. Comparison analysis curve of overlap rate and center position.



(a) Accuracy (b) Success rate

Fig. 11. Success rate and accuracy on the VOT dataset.



(a) Accuracy (b) Success rate

Fig. 12. Success rate and accuracy graph on the TrackingNet dataset.

Fig. 12 is a comparison chart of the accuracy and success rates of four algorithms on the tracking network dataset. The horizontal axis represents the position error threshold and overlap threshold, and the vertical axis represents the accuracy and success rate. The blue dashed line represents TTAAE, the red solid line represents TTFM, the green dashed line represents SiamRPN++, and the purple dashed line represents SiamFC. From Fig. 12 (a), it can be observed that as the position error threshold gradually increases, the four curves gradually decrease. Within a smaller range of position error thresholds, the TTAAF algorithm has the highest accuracy, followed by TTFM, then SiamRPN++, and finally SiamFC; When the position error threshold is greater than 0.3, the accuracy of SiamFC exceeds SiamRPN++. Fig. 12(b) shows the trend of success rate with respect to the overlap rate threshold. It can be seen that as the overlap rate threshold increases, the four curves show an upward and then downward trend. Within a larger range of overlap rate thresholds, the TTAAF algorithm has the highest success rate, followed by TTFM, then SiamRPN++, and finally SiamFC; when the overlap rate threshold is less than 0.3, the success rate of SiamFC exceeds that of SiamRPN++. Combining the two graphs, it can be concluded that the TTAAF algorithm has the best overall performance on the tracking network dataset, with an accuracy and success rate of 0.932 and 0.962, respectively, which are 0.3% and 0.5% higher than other algorithms. Especially when dealing with occlusion attribute

data, the performance of TTAAF algorithm is particularly outstanding, with accuracy and success rates of 0.642 and 0.753, respectively, which are 0.2% and 0.4% higher than other algorithms. This indicates that the TTAAF algorithm has good robustness in handling occlusion attribute data. Finally, it was applied to the actual target tracking process, and the experimental results are shown in Fig. 13.

From the experimental results in Fig. 13, it can be seen that there are differences in the performance of the target tracking algorithm in different scenarios. In scenes marked as' Normal ', the algorithm can successfully track targets, even in situations where there is occlusion or complex background between targets, as shown in the images in the upper and lower left corners, the algorithm can still accurately identify and track targets. However, in the scenario marked as' Tracking failed ', the tracking ability of the algorithm is challenged. The image in the upper right corner shows that when the target is partially obscured by the sofa, the algorithm cannot continue tracking the target, which may be due to confusion between the target features and background features, resulting in tracking loss. In the image in the lower right corner, the target also failed to track after entering the room due to changes in lighting and increased background complexity. This indicates that the algorithm may have limitations when dealing with lighting changes and complex backgrounds. In addition, the image on the right side of the middle shows that the algorithm can maintain tracking even when the target is moving rapidly, indicating that the algorithm has a certain robustness to dynamic scenes.

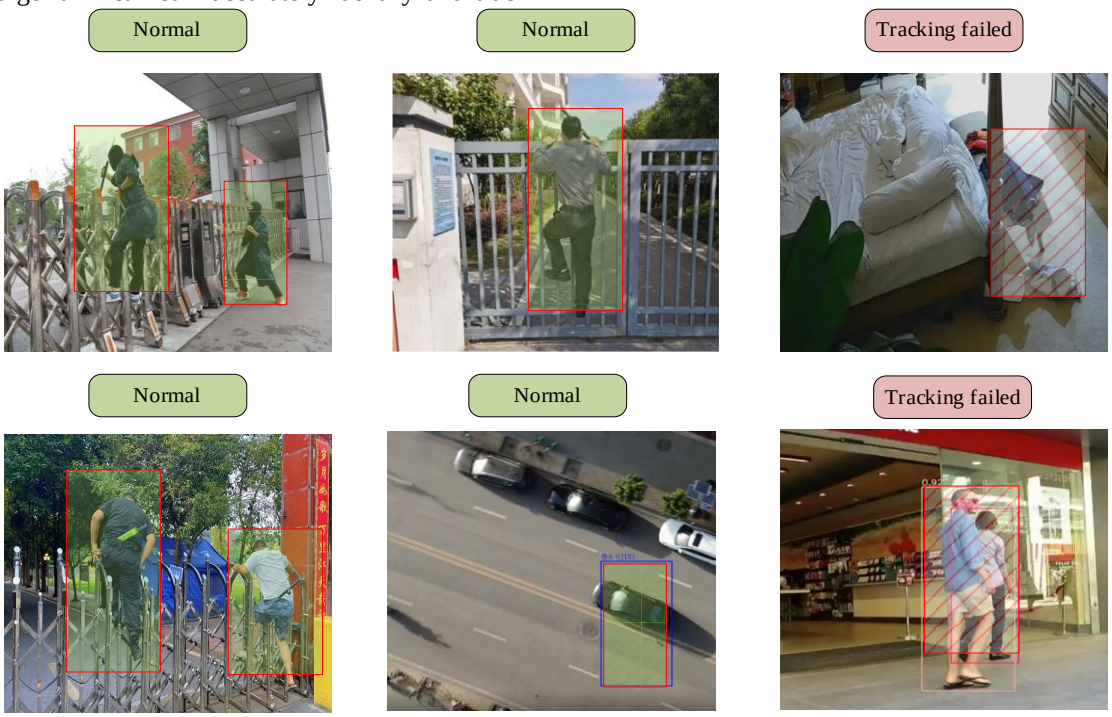


Fig. 13. Algorithm application effect interface diagram.

C. *Analysis of Twin Network Visual Tracking Algorithm Integrating Adaptive Anchor Box Generation*

A dual network visual tracking algorithm based on attention mechanism has been proposed to solve the problems in traditional fatigue driving detection, such as poor comfort, sensitivity to external factors, low detection accuracy, and poor real-time performance. By introducing attention mechanisms, algorithms can more effectively extract key features and improve target tracking performance in complex environments.

In the research of fatigue driving detection technology, this study proposes a visual tracking algorithm based on dual networks, which combines attention mechanism and adaptive anchor box generation. Compared with other methods in literature, it shows significant performance improvement. For example, the twin adaptive learning network proposed by Xin et al. performed well in GOT-10k and LaSOT benchmark tests, but the method proposed in this study can significantly reduce center position errors and improve tracking accuracy and robustness when dealing with complex background interference such as lighting changes, occlusion, and size changes [2]. The effective feature recognizer proposed by Jun W et al. optimized the representation ability, but there is still room for improvement in real-time performance [4]. In contrast, the overall success rate and accuracy of the algorithm in this study on the VOT dataset reached 0.831 and 0.862, respectively, which were 0.3% and 0.5% higher than other algorithms on average. Especially when processing occlusion attribute data, the success rate and accuracy were 0.2% and 0.4% higher than other algorithms on average, demonstrating better tracking performance and generalization ability.

The reasons and mechanisms for these results are as follows: the algorithm introduces attention mechanism, which can effectively filter out feature information that is not conducive to target tracking, retain key information, and improve feature extraction efficiency. Meanwhile, multi-layer linear fusion integrates feature information from different levels, enhancing feature richness and algorithm robustness. In addition, adaptive

anchor box generation reduces manual intervention and improves target tracking performance.

In summary, the dual network visual tracking algorithm based on attention mechanism proposed in this study has made significant progress in improving the accuracy and robustness of fatigue driving detection, providing an effective technical means for fatigue driving recognition in practical applications. Future research can further optimize algorithms to improve their adaptability and real-time performance in different driving scenarios, in order to better serve road safety and public health.

IV. CONCLUSION

Aiming at the challenges in fatigue driving detection, a visual tracking algorithm based on twin networks was proposed. Firstly, the TTFM algorithm was proposed. Then the TTAAF algorithm was optimized on the basis of the TTFM. The two algorithms proposed in the study performed well on multiple benchmark datasets, effectively improving the accuracy and robustness of target tracking. Specifically, the TTFM optimized the feature extraction process by introducing channel and spatial attention mechanisms, significantly enhancing the algorithm's ability to extract useful features. On the VOT dataset, the overall success rate and accuracy of the TTFM reached 0.504 and 0.642, with an average improvement of 0.6% and 0.9% compared with other algorithms. On the TrackingNet dataset, the success rate and accuracy of the TTFM algorithm were 0.542 and 0.763, especially for data containing occlusion attributes, which were on average 0.4% and 3.3% higher than other algorithms. Compared with TTFM, the TTAAF algorithm reduced manual intervention by adaptively generating anchor boxes, allowing the model to automatically learn and generate candidate boxes based on the semantic information of the image itself. On the VOT dataset, the overall success rate and accuracy of the TTAAF algorithm were 0.831 and 0.862, which were on average 0.3% and 0.5% higher than other algorithms. On the TrackingNet dataset, the TTAAF algorithm performed equally well, with an overall success rate and accuracy of 0.932 and 0.962, respectively, which were on average 0.3% and 0.5% higher than other algorithms. In summary, both algorithms proposed in the study have effectively improved the performance of target tracking, especially robustness in complex environments. This provides strong technical support for practical applications such as fatigue driving detection. Although significant progress has been made in this study, there are still some shortcomings. For example, the algorithms had high computational complexity and requires further optimization to improve real-time performance. Meanwhile, the robustness of the algorithm still needs to be improved for target tracking under extreme lighting conditions.

REFERENCES

[1] Xiaofeng B, Chenggang G. SiamMaskAttn: inverted residual attention block fusing multi-scale feature information for multitask visual object tracking networks. Signal, Image and Video Processing, 2023, 18 (2): 1305-1316.

[2] Bin P, Ke X, Ze'an L. Siamese refine polar mask prediction network for visual tracking. Signal, Image and Video Processing, 2023, 18 (1): 923-933.

[3] Xin L, Fusheng L, Wanqi Y. Siamada: visual tracking based on Siamese adaptive learning network. Neural Computing and Applications, 2024, 36 (14): 7639-7656.

[4] Jun W, Peng Y, Wenhui Y. Exploiting multi-scale hierarchical feature representation for visual tracking. Complex & Intelligent Systems, 2024, 10 (3): 3617-3632.

[5] Jianbin D. Design of smart operating table based on HCI and virtual visual tracking technology. International Journal on Interactive Design and Manufacturing (IJIDeM), 2023, 18 (2): 1019-1031.

[6] Di Y, Gu G, Xiu S. Self-supervised discriminative model prediction for visual tracking. Neural Computing and Applications, 2023, 36 (10): 5153-5164.

[7] Liu K, Liu L, Yang Sl. Spatial feature embedding for robust visual object tracking. IET Computer Vision, 2023, 18 (4): 540-556.

[8] Han W, Bo Z, Guizhong L. Refiner: a general object position refinement algorithm for visual tracking. Neural Computing and Applications, 2023, 36 (8): 3967-3981.

[9] Shaochuan Z, Tianyang X, Jun X W. A Spatio-Temporal Robust Tracker with Spatial-Channel Transformer and Jitter Suppression. International Journal of Computer Vision, 2023, 132 (5): 1645-1658.

[10] Chloe C, Isaac A M. Cognitive-perceptual traits associated with autism and schizotypy influence use of physics during predictive visual tracking. The European journal of neuroscience, 2023, 58 (10): 4236-4254.

[11] Zhang J, Yang X, Wang W. Cross-entropy-based adaptive fuzzy control for visual tracking of road cracks with unmanned mobile robot. Computer-Aided Civil and Infrastructure Engineering, 2023, 39 (6): 891-910.

[12] Yijin Y, Xiaodong G. Learning rich feature representation and aggregation for accurate visual tracking. Applied Intelligence, 2023, 53 (23): 28114-28132.

[13] Mengquan L, Xuedong W, Siming T. Visual tracking via confidence template updating spatial-temporal regularized correlation filters. Multimedia Tools and Applications, 2023, 83 (12): 37053-37072.

[14] Yuping Z, Zepeng Y, Bo M. Structural-appearance information fusion for visual tracking. The Visual Computer, 2023, 40 (5): 3103-3117.

[15] Liu Y, Jinchun D, Jun S. A new passive vision weld seam tracking method for FSW based on K-means. The International Journal of Advanced Manufacturing Technology, 2023, 128 (7-8): 3283-3295.

[16] An Z, Yi Z. Evota: an enhanced visual object tracking network with attention mechanism. Multimedia Tools and Applications, 2023, 83 (8): 24939-24960.

[17] Ning D, Kazuya T, Wenhui Jl. Estimation of control area in badminton doubles with pose information from top and back view drone videos. Multimedia Tools and Applications, 2023, 83 (8): 24777-24793.

[18] Amin S N, Shivakumara P, Jun T X, et al. An Augmented Reality-Based Approach for Designing Interactive Food Menu of Restaurant Using Android[C]//Artificial Intelligence and Applications. 2023, 1(1): 26-34.

[19] Nsugbe E. Toward a Self-Supervised Architecture for Semen Quality Prediction Using Environmental and Lifestyle Factors. Artificial Intelligence and Applications. 2023, 1(1): 35-42.

[20] Choudhuri S, Adeniye S, Sen A. Distribution Alignment Using Complement Entropy Objective and Adaptive Consensus-Based Label Refinement for Partial Domain Adaptation[C]//Artificial Intelligence and Applications. 2023, 1(1): 43-51.

Graph Neural Networks and Dominant Set Algorithms for Energy-Efficient Internet of Things Environments: A Review

Dezhi Liao*, Xueming Huang
School of Information Engineering, Huizhou Technician Institute, Huizhou, 516000, Guangdong, China.

***Abstract*—The widespread usage of Internet of Things (IoT) devices opens up new opportunities for automated operations, monitoring, and communications across various industries. However, extending the lifespan of IoT networks remains crucial because IoT devices are energy-limited. This study investigates the convergence of Graph Neural Networks (GNNs) and dominant set algorithms to extend the longevity of IoT networks. GNNs are neural networks that capture complex relationships and node interactions based on graph-structured data. With these capabilities, GNNs are extremely effective at modeling IoT network dynamics, where devices are connected and whose interactions have a significant impact on performance. In contrast, dominant set algorithms are defined as an approach in which nodes of a network function as agents or leaders to perform resource-efficient and resource-distributed communication. A further detailed overview leverages existing techniques to describe GNNs' role in optimizing dominant set algorithms and discusses integrating these technologies into addressing energy efficiency challenges in IoT settings.**

***Keywords*—Internet of Things; energy efficiency; dominant set; Graph Neural Networks**

I. Introduction

A. Context

In the last few years, technology has profoundly reached all areas of society and has transformed them. As technology has progressed during the last couple of shows, it has dramatically shifted our way of life and how we communicate, work, and live. Ultimately, all these lead to our lives based on data-driven decisions and greater connectedness. The digital transformation also benefits from the presence of potentially billions of connected objects, which now can be transformed into intelligent and connected ones through the Internet of Things (IoT) [1]. IoT encompasses a vast range of physical devices, from everyday items like refrigerators and thermostats to complicated equipment for industrial use [2]. These devices feature software, sensors, and a variety of components specifically engineered to capture and transmit data through the Internet [3].

The advancement of connectivity enhances the functionality and convenience of technologies [4]. The rapid growth of the Internet of Things (IoT) is fueled by advancements and widespread online access [5]. At the core of IoT is computing, which seamlessly integrates computers into our routines without us even noticing it. IoT design enables it to work discreetly in the background. Offers users a multitude of benefits and options. In industries and sectors today, IoTs are widely applied to show their flexibility and versatility [6]. One of the most relatable instances is how IoT transforms household tasks within residences, with the help of home technologies, making our daily lives more efficient and convenient.

B. Motivation

Energy efficiency is a critical concern in designing and operating IoT networks, given that most IoT devices operate on limited energy sources such as batteries, which are often difficult or impractical to replace [7]. The continuous operation of these devices, responsible for tasks ranging from environmental monitoring to smart city infrastructure, requires minimizing energy consumption to extend the network's life and ensure uninterrupted operations [8].

Typical ways to boost energy efficiency, like other algorithms such as duty cycling, energy-aware routing, and clustering algorithms, are often difficult to scale or adjust to changing network conditions. With the growth of IoT networks, these problems have become more apparent, and there is a need for new intelligent solutions to solve them. Significant energy savings can be achieved by reducing the number of active nodes through techniques such as dominant set algorithms and optimizing network operations using Graph Neural Networks (GNNs). These new ideas allow for quick changes to the network as they happen, cutting down on energy use and making IoT systems work better and last longer overall.

C. Contribution

The present study contributes to IoT network optimization. First, a comprehensive overview of the combination of GNNs and dominant set algorithms is provided, highlighting the potential of these ideas to enhance the energy efficiency of IoT networks. Second, existing methods are systematically reviewed, and several insights are provided regarding how GNNs are employed to optimize dominant sets and minimize energy consumption dynamically. Third, current challenges and limitations of traditional approaches are identified, and the synergistic use of GNNs and dominant set algorithms as a novel solution to these problems is proposed. Finally, future research directions are outlined, and the importance of scalability, real-

time adaptation, and security for developing more efficient and sustainable IoT networks is highlighted.

II. BACKGROUND

This section presents basic information on concepts and technologies that underpin the integration of GNNs and dominant set algorithms for energy efficiency in IoT networks.

A. IoT

The IoT, coined by Kevin Ashton in 1999, represents a global infrastructure in which physical and virtual entities are connected through advanced communications technologies. As defined by the International Telecommunication Union (ITU), IoT enables innovative services. As shown in Fig. 1, the IoT essentially represents a vast network of devices capable of collecting data, exchanging operational information, and performing autonomous tasks. Integrating sensors into various devices, from cell phones to home appliances, makes this functionality possible.

Fig. 1. IoT-related sectors.

An IoT ecosystem includes web-enabled devices with processors, sensors, and communications hardware to collect, transmit, and process data. Collected data is typically transferred to cloud platforms or processed locally before being shared with other connected devices to initiate actions [9]. While human interaction for configuration, guidance, or data access is still possible, IoT devices work independently of each other. IoT applications use specific connectivity, network, and communication protocols. In particular, IoT has the potential to leverage machine learning, a subset of artificial intelligence, to optimize data processing and improve system dynamics. IoT generates and analyzes enormous amounts of data in real-time, driving big data analytics. IoT enables companies to monitor employee performance across multiple locations and optimize operations.

B. Graph Neural Networks

GNNs represent a practical paradigm for processing graph-structured data, demonstrating exceptional performance across diverse domains, including physical systems, protein structure analysis, and knowledge graph management [10, 11]. Graphs can be classified based on edge directionality (directed or undirected), node and edge homogeneity (homogeneous or heterogeneous), and structural complexity (graphs or hypergraphs). Directed graphs exhibit unidirectional edges, while undirected graphs imply bidirectional relationships between connected nodes. Homogeneous graphs contain a single node and edge type, whereas heterogeneous graphs accommodate multiple types. Hypergraphs extend the graph concept by allowing edges to connect arbitrary numbers of vertices. Fig. 2 shows examples for each graph type up to this point.

A directed (simple) graph is formally defined as a tuple $G = (V, E)$, where V is a set of nodes and E is a set of directed edges represented as tuples. A directed (generalized) hypergraph is a similar tuple with hyperedges and a numbering map f_i for each edge to indicate node order. These graphs are considered elementary, as other graphs can be constructed from their composition. Directed graphs are undirected if $(u, v) \in \mathcal{E}$. In this case, edges can be represented as sets rather than tuples. Directed hypergraphs are undirected if $f_i: x \rightarrow \{0\} for\ all\ (x, f_i)_i \in \mathcal{E}$.

A multigraph is a graph where edges or nodes can appear multiple times. A heterogeneous graph is one where nodes or edges have different types. These types can be formally appended to nodes and edges. An attributed graph is one where nodes or edges are associated with attributes, represented by node and edge attribute functions. If only nodes have attributes, it's called node-attributed, and if only edges have attributes, it's called edge-attributed. The graph is considered weighted if edge attributes are a subset of edge types.

The core principle of GNNs is to iteratively aggregate information from neighboring nodes and integrate this aggregated data into the representation of the central node [12]. This process, known as spreading, repeats itself over several layers. Each layer includes aggregation and update operations formulated as follows.

$$Aggregation: n_v^{(l)} = Aggregator_l(\{h_u^l, \forall u \in N_v\})$$
$$Update: h_v^{(l|+1)} = Updater_l(h_u^l, n_v^{(l)}) \quad (1)$$

h_u^l represents the node u's representation at layer l, while $Aggregator_l$ and $Updater_l$ denote the respective functions for aggregation and update at layer l.

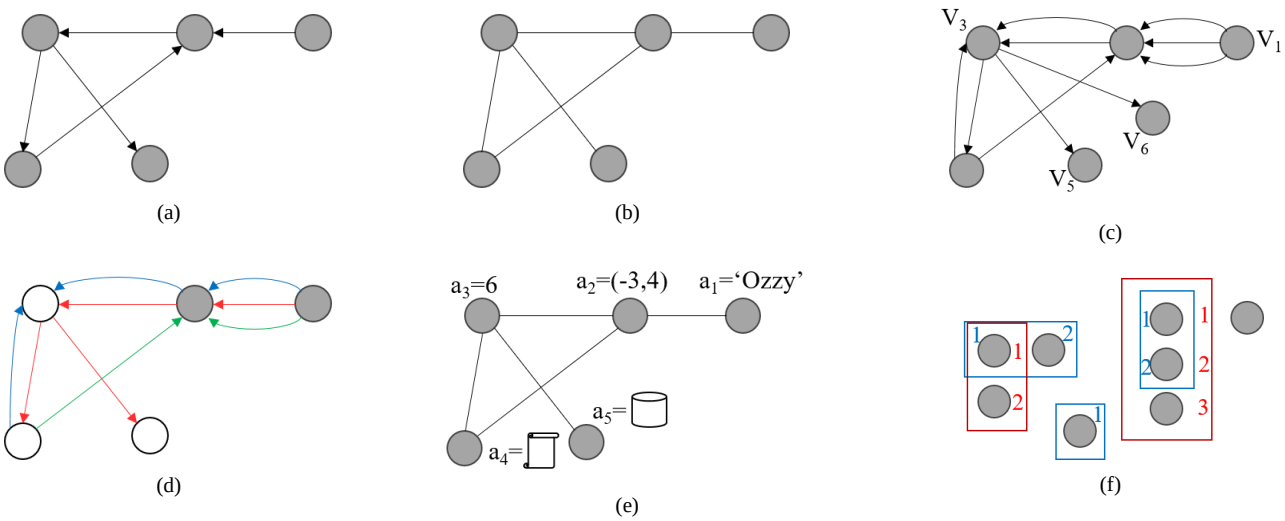


Fig. 2. An overview of different graph types: directed (a), undirected (b), multigraph (c), heterogeneous (d), attributed (e), and directed hypergraph (f).

TABLE I. COMPARISON OF AGGREGATION AND UPDATE OPERATIONS ACROSS DIFFERENT GNN FRAMEWORKS

Framework	Aggregation operation	Update operation
HGNN	Aggregates information using hyperedge convolution, where the node representations are transformed through a learnable matrix and normalized using degree matrices.	Updates node representations with a non-linear activation function applied to the aggregated information.
GGNN	Aggregates information uniformly from neighboring nodes using mean pooling.	Updates node representations using a gated recurrent unit, integrating the aggregated information iteratively.
GAT	Aggregates information from neighbors using an attention mechanism, where the contribution of each neighbor is weighted based on its relevance.	Updates node representations by applying a non-linear activation function to the weighted aggregation of neighbors' information.
GraphSAGE	Aggregates information from a sampled subset of neighbors using various pooling strategies (mean, sum, max).	Updates node representations by concatenating aggregated neighbor information with the underlying node representation and applying a learnable transformation.
GCN	Aggregates information using a weighted sum of neighbor representations, where the weights are derived from a normalized adjacency matrix.	Updates node representations by applying a non-linear activation function to the aggregated information.

Aggregation strategies include uniform treatment of neighbors (mean pooling) and weighted contributions based on attention mechanisms. The update step integrates the neighborhood information with the central node representation to create a refined node representation. Various techniques have been proposed to effectively combine these components, including GRU mechanisms, concatenation with non-linear transformations, and summation. Generally, five prominent GNN frameworks are commonly employed in IoT networks, each with distinct aggregation and update mechanisms, as summarized in Table I.

1) Hypergraph Neural Network (HGNN): This framework excels at capturing higher-order data correlations within hypergraph structures [13]. The hyperedge convolution layer defined by Eq. (2) uses a non-linear activation function, a learnable transformation matrix ($W^{(l)}$), and degree matrices for edges (D_e) and vertices (D_v) to calculate node representations.

$$Aggregation: N^{(l)} = D_v^{-\frac{1}{2}} E W^0 \tilde{D}_e^{-1} E^T D_v^{-\frac{1}{2}} H^{(l)}$$
$$Update: H^{(l+1)} = \delta\left(W^{(l)} N^{(l)}\right) \quad (2)$$

2) Gated Graph Neural Network (GGNN): GGNN incorporates a Gated Recurrent Unit (GRU) for the update step [14]. While effective, its iterative nature over all nodes can hinder scalability on large graphs.

$$Aggregation: n_v^{(l)} = \frac{1}{|N_v|} \sum_{j \in N_v} h_j^{(l)}$$
$$Update: h_v^{(l+1)} = GRU\left(h_v^{(l)}, n_v^{(l)}\right) \quad (3)$$

3) Graph Attention Network (GAT): Recognizing the varying influence of neighbors, GAT employs an attention mechanism to differentiate neighbor contributions [15]. The attention function, typically $LeakyReLU\left(a^T \left[W^{(l)} h_v^{(l)} \oplus W^{(l)} h_j^{(l)}\right]\right)$, assigns weights to neighbors based on their relevance.

$$Aggregation: n_v^{(l)} = \sum_{j \in N_v} a_{vj} h_j^{(l)}, a_{vj} = \frac{exp\left(Att\left(h_v^{(l)}, h_j^{(l)}\right)\right)}{\sum_{k \in N_v} exp\left(h_v^{(l)}, h_j^{(l)}\right)}$$
$$Update: h_v^{(l+1)} = \delta\left(W^{(l)} n_v^{(l)}\right) \quad (4)$$

4) GraphSAGE: This framework introduces neighborhood sampling to manage computational efficiency, followed by

aggregation (mean, sum, or max pooling) and concatenation for the update step [16].

$$Aggregation: n_v^{(l)} = Aggregator_l(\{h_u^l, \forall u \in N_v\})$$
$$Update: h_v^{(l+1)} = \delta\left(W^{(l)}.\left[h_v^{(l)} \oplus n_v^{(l)}\right]\right) \quad (5)$$

5) Graph Convolutional Network (GCN): GCN simplifies the aggregation process by approximating the graph Laplacian's eigendecomposition [17]The node representation is updated iteratively based on Eq. (6), which involves a non-linear activation function, a learnable transformation matrix, and adjacency weights.

$$Aggregation: n_v^{(l)} = \sum_{j \in N_v} d_{vv}^{-\frac{1}{2}} \tilde{a}_{vj} d_{jj}^{-\frac{1}{2}} h_j^{(l)}$$
$$Update: h_v^{(l+1)} = \delta\left(W^{(l)} n_v^{(l)}\right) \quad (6)$$

C. Dominant Set Algorithms

Dominant set algorithms identify subsets of nodes within a network capable of representing or leading a group of nodes, optimizing communication and resource allocation [18]. In the context of IoT networks, where energy efficiency is paramount, dominant set algorithms are crucial in minimizing the number of active nodes required for effective network operation [19]. By designating certain nodes as "dominant," these algorithms reduce overall communication overhead, conserving energy.

Let $G = (V, E)$ be a graph representing the IoT network, where $V = \{v_1, v_2, \ldots, v_n\}$ is the set of IoT nodes, and E is the set of edges representing the communication links between these nodes. A dominant set $D \subseteq V$ is a subset of nodes such that every node $v \in V$ is either in D or adjacent to at least one node in D. The goal is to minimize the size of the dominant set D while ensuring network coverage. This can be mathematically expressed as:

$$Minimize\ |D| \quad subject\ to \quad \forall v \in V, v \in D\ or\ \exists\ u \in D\ such\ that\ (u, v) \in E \quad (7)$$

In energy-constrained IoT networks, the energy consumption of a node v_i is denoted by $E(v_i)$. The total energy consumption of the dominant set D can be expressed as:

$$E(D) = \sum_{v_i} E(v_i) \quad (8)$$

The objective is to minimize $E(D)$ while maintaining a dominant set that covers the entire network, ensuring that:

$$Minimize\ E(D) \quad subject\ to \quad D\ is\ a\ dominant\ set\ of\ G \quad (9)$$

Traditional dominant set algorithms often rely on heuristic or optimization-based approaches to select the most suitable nodes. These algorithms typically consider node degree, connectivity, and proximity to other nodes. For instance, a common approach is to iteratively select the node with the highest degree (most connections) as part of the dominant set:

$$Select\ v_i \in V\ \ such\ that\ degree\ (v_i) \geq degree(v_j) \forall v_j \in V \quad (10)$$

This process is repeated until all nodes are either in the dominant set or adjacent to a node in the dominant set.

One of the main challenges of traditional dominant set algorithms is their static nature. They do not easily adapt to the dynamic conditions of IoT networks, such as fluctuating energy levels or changing network topologies. The computational complexity of finding the optimal dominant set can also be high, especially in large-scale IoT networks.

These limitations have spurred research into more advanced methods, such as integrating dominant set algorithms with machine learning techniques like GNNs. By leveraging the learning capabilities of GNNs, it is possible to develop more sophisticated dominant set algorithms that can dynamically adjust to real-time network conditions, offering a more robust solution for energy-efficient IoT network management. For example, GNNs can predict each node's energy consumption and connectivity dynamics, leading to a more effective and adaptive selection of the dominant set.

D. Energy Efficiency in IoT Networks

Energy efficiency is a paramount concern in the design and operation of IoT networks due to the inherent constraints of IoT devices, which often rely on limited power sources such as batteries. The operational lifespan of these devices and, by extension, the network depends heavily on how efficiently energy is utilized. Given that IoT networks are typically deployed in large numbers and diverse environments, ranging from smart cities to remote agricultural fields, the challenge of maintaining energy efficiency while ensuring continuous, reliable operation is critical. As shown in Fig. 3, the energy consumption in IoT networks is influenced by several factors, including:

1) Communication overhead: Data transmission and reception are among the most energy-intensive activities in IoT devices. The frequency of communication, the distance over which data must be transmitted, and the protocol used all significantly impact energy usage.

2) Idle listening: Nodes in an IoT network often consume energy while listening for potential communication, even if no data is transmitted. This idle listening can account for a substantial portion of energy expenditure.

3) Computation: Local processing tasks, such as data aggregation, encryption, or decision-making algorithms, consume energy. Although typically less than communication activities, computation energy must still be managed effectively, especially in devices with minimal processing power.

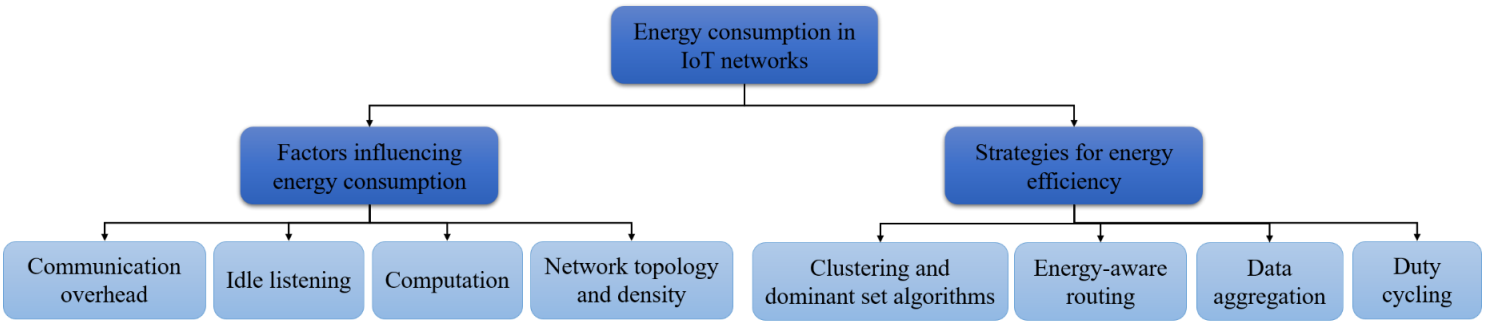


Fig. 3. Energy consumption in IoT networks.

4) Network topology and density: The arrangement and density of nodes within the network affect the routing paths and the number of hops required for data transmission, influencing energy consumption. Dense networks might reduce transmission distances but increase interference and collisions, while sparse networks might require longer transmission distances.

Various strategies have been developed to enhance energy efficiency in IoT networks, including:

1) Duty cycling: This technique involves turning off a device's radio transceiver when it is not needed, thus reducing energy consumption during idle periods. However, the challenge lies in coordinating wake-up times among nodes to maintain network connectivity.

2) Energy-aware routing: Routing protocols in IoT networks can be designed to consider the remaining energy of nodes when selecting routes, thus balancing energy consumption across the network and avoiding the depletion of individual nodes.

3) Clustering and dominant set algorithms: Clustering techniques group nodes into clusters with a designated cluster head, reducing the number of nodes involved in long-distance communication. As discussed in the previous section, dominant set algorithms are a form of clustering that selects a subset of nodes to manage communications, reducing energy consumption.

4) Data aggregation: Reducing the amount of data transmitted by aggregating or compressing data at intermediate nodes can significantly lower energy consumption. This strategy reduces the number of transmissions required and the volume of data sent [20].

Despite the effectiveness of these strategies, challenges remain, particularly in scaling these methods to large and heterogeneous networks, where nodes may have vastly different energy capacities, processing powers, and communication requirements. Furthermore, the dynamic nature of IoT networks, where nodes may move, join, or leave the network, complicates the implementation of static energy-saving techniques.

III. GNNs and IoT Networks

A. GNNs for Node Representation and Feature Learning

GNNs are effective tools for manipulating and examining graph-based data. This is especially useful for IoT networks, in which devices (nodes) and their connections (edges) naturally form graphs. One of the main advantages of GNNs is their capacity to acquire efficient node representations and extract significant characteristics from the graph. These features may be used for various downstream tasks like optimization, prediction, and classification.

Within IoT systems, individual nodes symbolize devices with distinct features, like energy levels, communication range, computing power, and connectedness to other nodes. The connections between nodes, such as the ability to directly interact with one another, are represented as edges in the graph. Conventional approaches to representing nodes typically depend on predetermined characteristics or rules, which may not completely portray the many relationships and interdependencies in a changing network environment.

GNNs address this constraint by acquiring node representations influenced by the overall network topology. This is accomplished by passing a message in which each node continuously changes its representation by gathering information from neighboring nodes. The whole procedure may be represented numerically in the following manner [21].

$$h_v^{(k)} = \sigma\left(W^{(k)}.AGGREGATE\left(\left\{h_u^{(k-1)}: u \in N(v)\right\}\right)\right) \quad (11)$$

Where $h_v^{(k)}$ is the representation of node v at the k^{th} iteration, $N(v)$ denotes the set of neighbors of node v, *AGGREGATE* is a function that combines the representations of neighboring nodes (e.g., sum, mean, max), $W^{(k)}$ is a learnable weight matrix for the k^{th} layer, and σ is a non-linear activation function (e.g., ReLU).

During this repeated process, the encoding of each node $h_v^{(k)}$ includes the node's characteristics, the characteristics of its neighboring nodes, and the wider network context. GNNs can acquire comprehensive and contextually aware representations of nodes, which are very useful for tasks like clustering, dominant set selection, and energy-efficient routing in IoT

networks. GNNs may capture many crucial components of the network via the learned features, including:

1) Node centrality: This feature denotes the significance or impact of a node inside the network.

2) Connectivity patterns: It comprehensively represents local and global network configuration, including identifying nodes that form clusters or key routes.

3) Energy consumption patterns: It uses predictive analysis to identify nodes more likely to spend more energy depending on their position in the network. This allows for proactive load balancing.

GNNs may enhance decision-making processes in IoT networks by using these acquired characteristics. In the dominant set method, nodes with high centrality and low energy consumption are prioritized as leaders. This optimization aims to improve network coverage and energy efficiency. Utilizing GNNs for node representation and feature learning in IoT networks offers the following benefits:

1) Scalability: GNNs can effectively manage networks of significant size by gaining knowledge in a decentralized and concurrent fashion.

2) Adaptability: The learned representations may be continuously modified as the network grows, enabling immediate adjustment to changes in network structure or node conditions.

3) Generalization: GNNs can apply learned knowledge to diverse network setups, which allows them to remain strong and effective even when faced with varied situations and network architectures.

B. GNNs for Clustering and Dominant Set Selection

GNNs have shown substantial promise in improving IoT network clustering and dominant set selection procedures. Clustering represents categorizing nodes (IoT devices) into smaller groups, known as subsets or clusters. Each cluster is led by a cluster head, who handles communication and coordination inside the cluster. Dominant set selection is a notion that involves choosing a subset of nodes to serve as leaders or representatives responsible for controlling communication throughout the whole network. These processes are essential, enhancing energy efficiency, lowering communication overhead, and lengthening IoT networks' lifespan.

Conventional clustering methods often depend on fixed rules or predetermined standards, which may not completely adjust to the dynamic nature of IoT environments. GNNs provide a more adaptable and data-oriented clustering method by using acquired representations of nodes that include both the local and global structure of the network.

Within the realm of clustering, GNNs generate embeddings reflecting the similarity between nodes, taking into account their characteristics and connections. These embeddings may be used to create clusters more informedly. Nodes that possess comparable energy levels, communication patterns, or responsibilities within the network may be categorized together, establishing more efficient clusters. Mathematically, after learning node representations h_v through a GNN, clustering can be performed by applying a clustering algorithm, such as k-means, directly on these embeddings [21].

$$Cluster\ assignment = k - means(\{h_v | v \in V\}) \quad (12)$$

This methodology guarantees the formation of clusters based on acquired, multi-dimensional characteristics that include intricate interconnections and dependencies within the network instead of merely depending on unprocessed, pre-established criteria.

Dominant set selection entails identifying a subset of nodes that can efficiently handle communication and data aggregation for the whole network. Conventional approaches for choosing dominating sets often include picking nodes based on node degree, closeness, or energy levels. Nevertheless, these approaches may be constrained by their fixed characteristics and incapacity to adjust to evolving network circumstances.

GNNs can be used to dynamically improve the selection of dominant sets by learning to predict the most suitable nodes as representatives. This prediction considers both the network's current state and likely future situations. The method can be defined as a node classification problem in which the GNN is trained to categorize nodes into dominant (leader) or non-dominant (follower) groups based on their acquired embeddings [21].

$$y_v = GNN(G, v) \quad (13)$$

where, y_v is the predicted label for node *v*, indicating whether it should be included in the dominant set. The GNN model considers the entire graph *G* and the specific node *v*, using the message-passing framework to aggregate information from neighboring nodes and the broader network. The inclusion of a node in the dominant set may be determined by a range of acquired characteristics, including:

1) Energy efficiency: Nodes with more remaining energy may be prioritized to prevent the depletion of crucial nodes.

2) Network centrality: Nodes centrally positioned or with excellent connections are more suitable since they can effectively interact with other nodes.

3) Load balancing: The GNN may spread the predominant function across numerous nodes to prevent certain nodes from becoming bottlenecks.

The use of GNNs in clustering and dominate set selection has noteworthy ramifications for IoT networks, especially when conserving energy and ensuring network lifespan is of utmost importance.

1) Smart cities: Within urban IoT networks, GNNs can enhance the arrangement of sensors and devices into clusters, decreasing energy use while ensuring a reliable connection and efficient data processing.

2) Industrial IoT: In industrial environments, GNNs may effectively oversee the management of crucial monitoring devices, reducing energy depletion in vital nodes.

3) Environmental monitoring: For IoT networks deployed in remote or difficult-to-access areas, GNNs can help form optimal clusters and select dominant sets that minimize energy usage, extending the network's operational lifespan.

IV. Dominant Set Algorithms in IoT Networks

A. Traditional Dominant Set Algorithms

Traditional dominant set methods were previously employed in IoT networks to enhance network efficiency and minimize energy use by choosing a subset of nodes that efficiently handle communication and data aggregation duties. These algorithms identify a "dominant set" of nodes that may include the whole network, guaranteeing that every node is either a member of the dominant set or is directly linked to a node in the dominant set. This method effectively minimizes the required communication since it only involves a small number of critical nodes in transmitting messages and handling data. This saves energy across the network.

The greedy algorithm is a frequently used method for identifying a dominant set. It operates by repeatedly selecting nodes with the greatest degree (i.e., the nodes with the most connections) to be included in the dominant set. The reasoning is that nodes with a high degree are more likely to include a significant percentage of the network, decreasing the number of nodes required in the dominating set. The fundamental procedures of a greedy dominating set algorithm may be summarized as follows:

Start with an empty set $D = \emptyset$.

Select the node v with the maximum degree in the remaining graph and add it to D.

Remove v and its neighbors from the graph.

Repeat until all nodes are either in D or have a neighbor in D.

Heuristic techniques are often used to enhance the efficiency of selecting dominating sets. These approaches may include other factors such as node energy levels, proximity to other nodes, or the general structure of the network. For instance, a heuristic may prioritize nodes with more energy to prevent the dominant set from rapidly exhausting its resources.

Computing the precise minimal dominating set in large-scale IoT networks may be computationally burdensome. Approximation algorithms provide a means to discover solutions close to optimum efficiently. These algorithms generally ensure that the size of the dominating set is within a certain ratio of the ideal size. An example of a well-recognized approximation technique is the 2-approximation algorithm. This approach ensures that the dominating set it discovers will be no more than twice the size of the best answer.

B. Challenges in Existing Approaches

Although conventional dominating set algorithms have played a crucial role in enhancing the energy efficiency of IoT networks, they still face some difficulties, as outlined in Table II. With IoT networks' growing complexity and dynamism, certain limitations in current techniques have been revealed. These problems highlight the need for more sophisticated approaches to effectively manage contemporary IoT systems' distinct requirements.

A major obstacle classic dominating set algorithms face is their dependence on static network assumptions. These methods typically function using a static snapshot of the network, assuming that the network's architecture, node connection, and energy levels stay unchanged throughout the network's operation. Nevertheless, IoT networks include an intrinsic dynamism, where nodes can join or depart from the network, relocate to new positions, or encounter variations in energy levels. Using a static technique may lead to the formation of inefficient dominating sets that do not adjust to the changing circumstances of the network. This can result in inefficiencies in both energy usage and communication.

TABLE II. Challenges of Conventional Dominant Set Algorithms in IoT Networks

Challenge	Description	Impact on IoT Networks
Static network assumptions	Dependence on a static snapshot of the network, assuming fixed architecture, node connections, and energy levels.	Leads to inefficient dominating sets that do not adapt to dynamic network changes, resulting in energy inefficiencies and communication issues.
Scalability issues	Increasing computational complexity as the network size grows, making it difficult to find efficient dominating sets in large-scale networks.	Limits the applicability of conventional methods in large-scale IoT networks, causing performance bottlenecks and reducing overall network efficiency.
Energy unawareness	Focus on network coverage and connectivity without considering the varying energy levels of individual nodes.	Causes rapid depletion of selected nodes' energy, leading to uneven energy distribution and shortening the overall network lifespan.
Lack of real-time adaptability	Inability to adapt to real-time changes in the network, such as node mobility or environmental fluctuations.	Results in the use of outdated configurations, leading to inefficiencies in energy use and increased energy consumption.
Suboptimal node selection	Reliance on simple heuristics like node degree or proximity, which may not consider the complex factors affecting network performance.	May result in the selection of nodes that are not optimally suited to act as leaders, reducing the network's overall efficiency and effectiveness.
Security vulnerabilities	Algorithms not designed with security in mind, making them susceptible to attacks like node compromise or denial-of-service (DoS) attacks.	Leaves IoT networks vulnerable to security breaches, as crucial nodes in the dominating set may be targeted by adversaries, disrupting network operations.

As the size of IoT networks grows, the computational intricacy of conventional dominating set methods becomes a notable issue. Discovering the most efficient or nearly efficient dominating set in a vast network may be a demanding and time-consuming computing task. The complexity of these algorithms increases exponentially as the number of nodes in the network rises, making it unfeasible to apply them to large-scale IoT networks without experiencing significant performance limitations. The limited scalability of conventional techniques hinders their usefulness in contexts with broad and constantly growing IoT networks.

Traditional dominant set algorithms often focus on optimizing network coverage and connectivity without fully considering the varying energy levels of individual nodes. In many cases, these algorithms may select nodes with low remaining energy to be part of the dominant set, leading to rapid depletion of those nodes' power reserves. This "energy blindness" can result in uneven energy distribution across the network, with some nodes exhausting their energy supply quickly while others remain underutilized. Consequently, the overall network lifespan may be shortened as key nodes fail prematurely due to energy depletion.

Another notable constraint is the lack of real-time adaptability of conventional dominant set methods to network changes. IoT networks often function in dynamic settings characterized by fast changes in circumstances, such as mobile IoT scenarios or networks installed in harsh and fluctuating environments. Conventional algorithms, usually created to calculate a dominating set by a single network examination, have difficulty keeping up with these modifications. Consequently, they could persist in using obsolete settings, resulting in inefficiencies and heightened energy use.

Classic dominant set methods often rely on node degree, proximity, or simple heuristics to select a node. However, these criteria may not necessarily result in the most energy-efficient or effective dominant set. These criteria sometimes fail to include the intricate and multi-faceted elements that impact the functioning of an IoT network, such as the diverse responsibilities of nodes, environmental circumstances, or the unique communication patterns inside the network. This may lead to the selection of nodes not optimally suited to act as leaders, reducing the network's overall efficiency.

Security is a crucial problem in several IoT applications. Conventional dominating set algorithms, on the other hand, are often not developed with security as a primary consideration. The algorithms' static and predictable nature renders them susceptible to assaults, such as node compromise or denial-of-service (DoS) attacks. In these attacks, an adversary targets crucial nodes in the dominating set to disrupt the network. Conventional methods might leave IoT networks vulnerable to possible security breaches without inherent mechanisms to adjust to such threats.

C. Enhancements through GNN Integration

Incorporating GNNs with conventional dominant set techniques signifies notable progress in overcoming the constraints of current methodologies in IoT networks. As summarized in Table III, GNNs can effectively represent and analyze intricate data structures through graphs. This makes them a powerful tool for improving the process of selecting dominating sets and clustering in IoT settings. GNNs provide a flexible, scalable, and adaptable framework for this purpose. This part examines the integration of GNNs with dominant set algorithms to address the issues related to static assumptions, scalability, energy efficiency, and real-time adaptation.

TABLE III. ENHANCEMENTS PROVIDED BY GNN INTEGRATION WITH DOMINANT SET ALGORITHMS IN IOT NETWORKS

Enhancement	Description	Impact on IoT Networks
Dynamic network adaptation	GNNs continuously update node representations in real-time as network conditions evolve, ensuring the dominant set adapts to current network states.	Maintains optimal performance and energy efficiency by dynamically responding to node mobility, energy changes, and communication patterns.
Scalability and efficiency	GNNs handle large-scale IoT networks efficiently through parallel processing and distributed computing, minimizing global calculations.	Enhances scalability of dominant set algorithms, making them suitable for large-scale deployments like smart cities and industrial IoT systems.
Energy-aware node selection	GNNs incorporate energy-awareness into node selection by analyzing current and historical energy data, prioritizing nodes with higher remaining energy.	Extends network lifetime by preventing the premature depletion of key nodes and ensuring balanced energy consumption across the network.
Improved node representation	GNNs generate rich, context-aware representations of nodes that include their characteristics and network context, enabling more informed dominant set selection.	Optimizes energy consumption and maintains network coverage by selecting strategically positioned, well-connected nodes for the dominant set.
Real-time decision making	GNNs support continuous updates and real-time decision-making in volatile network conditions, allowing for ongoing adjustments to the dominant set.	Ensures the network remains efficient and resilient even in dynamic environments like mobile IoT networks or uncertain deployment areas.
Security and robustness	GNNs enhance network security by detecting and responding to abnormal patterns, such as compromised nodes, and adjusting the dominant set accordingly.	Increases the resilience of IoT networks against security threats by proactively isolating or bypassing suspicious nodes, maintaining overall network integrity.

GNNs provide a significant improvement by being able to adjust to changing network circumstances in real-time. Contrary to conventional dominant set techniques that usually function with fixed network snapshots, GNNs continuously modify node representations as network circumstances evolve. GNNs accomplish this dynamic adaptation by using the message-passing process, which involves iteratively updating the representation of each node depending on the information it receives from its neighboring nodes. GNNs can adapt the dominating set to match the current state of the network, which includes factors such as node mobility, changes in energy levels, and different communication patterns. This ensures that GNNs can maintain maximum performance and energy efficiency as the network develops.

GNNs have an innate ability to scale, which makes them highly suitable for accommodating large-scale IoT networks. Conventional dominant set techniques sometimes encounter difficulties due to the high computing cost of finding the most effective nodes within extensive networks. GNNs, on the other hand, are capable of effectively managing enormous graphs via the use of parallel processing and distributed computing. By acquiring data from local neighborhoods in the network, GNNs minimize the need for global calculations, enabling scalable solutions as the network expands. The capacity to scale allows dominating set algorithms to be improved with GNN in large-scale IoT deployments, such as smart cities or industrial IoT systems, where conventional methods may struggle.

One significant benefit of combining GNNs with dominant set algorithms is the capacity to include energy-awareness in selecting nodes. GNNs can analyze node information, such as current energy levels, historical energy use, and connection, to learn and anticipate trends in energy consumption throughout the network. This data may be used to prioritize nodes with more remaining energy or to divide the communication workload equally across the network, thereby preventing the premature exhaustion of crucial nodes. Using GNNs for node selection, focusing on energy efficiency, extends the network's lifetime. This is achieved by preventing any one node from being excessively burdened, resulting in a more equitable distribution of energy consumption.

GNNs are very efficient in acquiring comprehensive and contextually aware representations of individual nodes within a network. The acquired representations include not only the immediate characteristics of the nodes (such as energy levels and connections) and the wider structural environment in which these nodes function. The GNN's learned embeddings include several aspects of a node, such as its centrality, its position in the network, and its closeness to other important nodes. These improved representations enable a more advanced and knowledgeable selection of dominating sets, where the chosen nodes are well-connected and strategically positioned to optimize energy consumption and sustain network coverage.

GNNs can efficiently handle and acquire knowledge from constantly changing data, making them especially powerful when network circumstances are volatile. In scenarios like mobile IoT networks or networks deployed in uncertain areas, GNNs can consistently update the dominating set as nodes relocate or as environmental conditions change. This real-time decision-making capability guarantees the network maintains its efficiency and resilience, even when faced with continuous changes. GNNs may be used with reinforcement learning methods to improve their capacity to adapt and optimize over time, leading to ongoing improvements in the performance of the dominant set algorithm.

By incorporating GNNs into dominant set algorithms, the security and resilience of IoT networks are improved. GNNs may be taught to detect and react to abnormal patterns in a network that may suggest security risks, such as nodes that have been hacked or strange traffic patterns. GNNs may safeguard the network from assaults and preserve its general operation by adapting the dominating set to bypass or isolate suspect nodes. The proactive strategy towards security, together with the robust adaptability of GNNs in managing network dynamics, leads to a more resilient IoT network that can survive various problems.

V. Conclusion

The exponential expansion of the IoT has introduced fresh prospects for automation, surveillance, and communication in several fields. Nevertheless, the energy efficiency of these networks continues to be a significant obstacle because of the constrained power resources of IoT devices. This research has examined the combination of GNNs with dominant set algorithms as a new method to improve the energy efficiency and operational longevity of IoT networks. We have examined conventional dominating set algorithms, emphasizing their use in minimizing communication overhead and preserving energy in IoT networks. Nevertheless, these approaches encounter substantial obstacles, such as their fixed characteristics, problems with expanding to larger scales, insufficient capacity to adjust in real-time, and inadequate consideration of energy fluctuations among nodes. To overcome these restrictions, incorporating GNNs presents a favorable option, as it brings about dynamic adaptability, scalability, energy-conscious node selection, and enhanced node representation, all of which are crucial aspects of IoT network management.

The GNN-based advances outlined in this research, including dynamic network adaptability, real-time decision-making, and security upgrades, signify significant progress compared to conventional methods. These improvements enable IoT networks to be more robust, streamlined, and capable of managing the intricacies of contemporary IoT settings. Nevertheless, there are still obstacles to overcome, such as the need for GNN models that can be easily expanded and operate well, the implementation of strong security and

privacy measures, and the meticulous handling of trade-offs between energy efficiency and other factors. To fully harness the capabilities of GNN-based dominant set algorithms in IoT networks, it is essential to prioritize ongoing research and development in this field. To construct more intelligent and adaptable IoT networks that match the expectations of future applications, we may solve difficulties and explore new avenues, such as integrating GNNs with other optimization methods. The combination of GNNs and dominant set algorithms provides a robust foundation for developing IoT networks that are both sustainable and energy-efficient, enabling them to operate well in dynamic and resource-limited contexts.

REFERENCES

[1] B. Pourghebleh and V. Hayyolalam, "A comprehensive and systematic review of the load balancing mechanisms in the Internet of Things," Cluster Computing, pp. 1-21, 2019.

[2] O. A. Ilori, A. A. Willoughby, O. F. Dairo, and A. O. Soge, "Design and Construction of a Photovoltaic Monitoring System Based on Wireless Sensor Networks and Internet of Things Technology," Journal of The Institution of Engineers (India): Series B, pp. 1-16, 2024.

[3] F. Kamalov, B. Pourghebleh, M. Gheisari, Y. Liu, and S. Moussa, "Internet of medical things privacy and security: Challenges, solutions, and future trends from a new perspective," Sustainability, vol. 15, no. 4, p. 3317, 2023.

[4] A. Mukherjee, K. Chatterjee, P. K. Kundu, and A. Das, "Probabilistic neural network-aided fast classification of transmission line faults using differencing of current signal," Journal of The Institution of Engineers (India): Series B, pp. 1-14, 2021.

[5] Z. Lu and X. Deng, "A cloud and IoT-enabled workload-aware Healthcare Framework using ant colony optimization algorithm," International Journal of Advanced Computer Science and Applications, vol. 14, no. 3, 2023.

[6] L. Hu and Y. Shu, "Enhancing Decision-Making with Data Science in the Internet of Things Environments," International Journal of Advanced Computer Science and Applications, vol. 14, no. 9, 2023.

[7] V. Moudgil, K. Hewage, S. A. Hussain, and R. Sadiq, "Integration of IoT in building energy infrastructure: A critical review on challenges and solutions," Renewable and Sustainable Energy Reviews, vol. 174, p. 113121, 2023.

[8] M. H. Alsharif et al., "A comprehensive survey of energy-efficient computing to enable sustainable massive IoT networks," Alexandria Engineering Journal, vol. 91, pp. 12-29, 2024.

[9] B. Pourghebleh, N. Hekmati, Z. Davoudnia, and M. Sadeghi, "A roadmap towards energy - efficient data fusion methods in the Internet of Things," Concurrency and Computation: Practice and Experience, vol. 34, no. 15, p. e6959, 2022.

[10] G. Corso, H. Stark, S. Jegelka, T. Jaakkola, and R. Barzilay, "Graph neural networks," Nature Reviews Methods Primers, vol. 4, no. 1, p. 17, 2024.

[11] B. Khemani, S. Patil, K. Kotecha, and S. Tanwar, "A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions," Journal of Big Data, vol. 11, no. 1, p. 18, 2024.

[12] R. Liu, P. Xing, Z. Deng, A. Li, C. Guan, and H. Yu, "Federated graph neural networks: Overview, techniques, and challenges," IEEE Transactions on Neural Networks and Learning Systems, 2024.

[13] N. Yin et al., "Messages are never propagated alone: Collaborative hypergraph neural network for time-series forecasting," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.

[14] C. Liu, Y. Li, H. Lin, and C. Zhang, "GNNRec: Gated graph neural network for session-based social recommendation model," Journal of Intelligent Information Systems, vol. 60, no. 1, pp. 137-156, 2023.

[15] U. A. Bhatti et al., "MFFCG–Multi feature fusion for hyperspectral image classification using graph attention network," Expert Systems with Applications, vol. 229, p. 120496, 2023.

[16] T. Liu, A. Jiang, J. Zhou, M. Li, and H. K. Kwan, "GraphSAGE-based dynamic spatial–temporal graph convolutional network for traffic prediction," IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 10, pp. 11210-11224, 2023.

[17] H. T. Phan, N. T. Nguyen, and D. Hwang, "Aspect-level sentiment analysis: A survey of graph convolutional network methods," Information Fusion, vol. 91, pp. 149-172, 2023.

[18] W. Zheng and B. Doerr, "Mathematical runtime analysis for the non-dominated sorting genetic algorithm II (NSGA-II)," Artificial Intelligence, vol. 325, p. 104016, 2023.

[19] B. Alwasel, A. Salim, A. M. Khedr, and W. Osamy, "Dominating sets-based approach for maximizing Lifetime of IoT-based heterogeneous WSNs enabled sustainable smart city applications," IEEE Access, vol. 12, pp. 44069-44079, 2024.

[20] B. Pourghebleh and N. J. Navimipour, "Data aggregation mechanisms in the Internet of things: A systematic review of the literature and recommendations for future research," Journal of Network and Computer Applications, vol. 97, pp. 23-34, 2017.

[21] J. M. Thomas, A. Moallemy-Oureh, S. Beddar-Wiesing, and C. Holzhüter, "Graph neural networks designed for different graph types: A survey," arXiv preprint arXiv:2204.03080, 2022.

Development of Traffic Light and Road Sign Detection and Recognition Using Deep Learning

Towards Safe and Robust Sensor-Perception System of Autonomous Vehicle Development Research

Joseph M. De Guia[1], Madhavi Deveraj[2]
School of Information Technology (SOIT), Mapua University, Manila, Phillippines[1, 2]
Energy Research Institute (ERI@N), Nanyang Technological University, Singapore[1]

***Abstract*—Traffic light and road sign violations significantly contribute to traffic accidents, particularly at intersections in high-density urban areas. To address these challenges, this research focuses on enhancing the accuracy, robustness, and reliability of Autonomous Vehicle (AV) perception systems using advanced deep learning techniques. The novelty of this study lies in the comprehensive development and evaluation of real-time traffic light and road sign detection systems, comparing state-of-the-art models including YOLOv3, YOLOv5, and YOLOv7. The models were rigorously tested in a controlled offline environment using the Nvidia Titan RTX, followed by extensive field testing on an AV test vehicle equipped with sensor suite and Nvidia RTX GPU. The testing was conducted across complex urban driving scenarios at the CETRAN proving test track, JTC Cleantech Park, and NTU Singapore campus. The traffic light detection and recognition (TLR) results demonstrate that YOLOv7 outperforms YOLOv5 and YOLOv3, achieving a mean Average Precision (mAP@0.5) of 93%, even under challenging conditions like poor lighting and occlusions. While the traffic road sign detection (TSD) mAP@0.5 of 96%. This superior performance highlights the potential of YOLOv7 in enhancing AV safety and reliability. The conclusions underscore the effectiveness of YOLOv7 for real-time detection in AV perception systems, offering crucial insights for future research. Potential implications include the development of more robust and accurate AV systems, capable of safely navigating complex urban environments.**

***Keywords*—Artificial intelligence; autonomous vehicle; traffic light recognition; road sign detection; YOLO; real-time object detection**

I. INTRODUCTION

Traffic light and road sign violations are a significant concern in Singapore, contributing to the rising number of traffic accidents, particularly at intersections. In 2023, over 31,815 red-light running violations were recorded, leading to numerous collisions and injuries at intersections [1]. Reports from the Singapore Police Force and Channel NewsAsia highlight the growing trend of drivers running red lights, resulting in dangerous collisions and fatalities [2]. The WHO Global Status Report on Road Safety 2023 also emphasizes that traffic signal violations are a major cause of road traffic injuries and fatalities globally [3]. These violations are especially problematic in high-density areas like the Central Business District (CBD) and busy residential and school campus zones, where the volume of both vehicular and pedestrian traffic is high. The failure to obey traffic signals in these areas can lead to hazardous situations, such as collisions with pedestrians at crossings or crashes involving multiple vehicles at junctions.

To address these challenges, recent research in AV development has focused on the integration of advanced perception systems that can accurately detect and classify traffic lights [4] and road signs in real-time [5]. These systems, leveraging deep learning algorithms and high-precision sensors, are designed to operate effectively even under challenging conditions such as poor lighting or heavy rain. By ensuring that AVs can reliably recognize and respond to traffic lights and signs, these technologies hold the potential to significantly reduce traffic violations, prevent accidents, and enhance overall road safety. Refer to Fig. 1 for the demonstration of the detection capability of the proposed TLR and TSD system.

(a) (b)

Fig. 1. Traffic light and road sign detection results. (a) The system accurately identifies a green light (95% confidence) and a red right arrow (94% confidence) at night. (b) The system detects a zebra crossing sign with 85% confidence in daylight. These results demonstrate the system's ability to detect and classify signals and signs across varying lighting conditions, essential for AV safety.

Previous studies have explored various approaches to traffic light and road sign detection, including traditional image processing techniques, machine learning models, and more recently, deep learning methods. While traditional methods offered limited success due to their inability to generalize across different environments and lighting conditions, deep learning has emerged as a powerful tool to improve the accuracy and robustness of real-time object detection and classification [4][5].

Notable advancements include the use of convolutional neural networks (CNNs) in models such as Fast R-CNN [6], SSD [7], and YOLO (You Only Look Once) [8], which have demonstrated superior performance in detecting a wide range of objects in images.

However, despite these advancements, challenges remain in achieving high recognition accuracy and stability under diverse and dynamic environmental conditions. Environmental variability, including changing lighting and weather conditions, as well as occlusions by other objects, can hinder accurate detection [9]. Real-time processing constraints, such as the need for high computational efficiency and low latency, further complicate the deployment of these systems [10]. Additionally, generalizing detection models across diverse geographical regions, handling data annotation and training challenges, and ensuring robustness against adversarial attacks and physical manipulation are ongoing issues [11].

Considering these challenges, this research focuses on advancing real-time detection and recognition of traffic lights and road signs using state-of-the-art deep learning techniques. By exploring the performance of different YOLO variants - YOLOv3 [12], YOLOv5 [13], and YOLOv7 [14], this study aims to improve the recognition accuracy and stability of AV perception systems. The research is particularly focused on high-risk areas such as junctions and pedestrian crossings, where the need for precise and reliable detection is paramount.

The novelty of this work lies in its comprehensive approach to evaluating and improving AV perception systems under real-world conditions. By conducting extensive testing in complex urban environments, this research contributes to filling critical gaps in current AV technologies. The study's objectives include the development and deployment of an optimal model for traffic light detection and recognition (TLR) and traffic road sign detection (TSD), evaluating system performance to enhance accuracy and reliability, and reporting the results of real-world testing. Our threefold contribution to the research:

- Performance Evaluation and Enhancement: The study conducted comprehensive evaluations of the perception system, leading to improvements in detection accuracy, robustness, and reliability, particularly under challenging environmental conditions.
- Comparative Analysis of YOLO Variants: The research compared the performance of multiple YOLO variants - YOLOv3, YOLOv5, YOLOv7 in real-time detection tasks, providing insights into their strengths and weaknesses in the context of AV perception systems.
- Real-World Testing and Validation: The research involved testing, verification, validation, and self-assessment of the deployed model in real-world test environments, including diverse urban scenarios in Singapore.
    - Ultimately, this research aims to contribute to the advancement of AV technologies, ensuring their safe and effective integration into public roadways.
    - The paper is divided into six sections. Section II describes the related works to TLR and TSD recognition research. In Section III we present in details of the TLR and TSD perception development in the AV research platform. Section IV is the Methodology describing our design, development, and evaluation of TLR and TSD systems in the AV research platform sensor and perception systems using custom-trained YOLO variant models. The results and discussion of the evaluation and assessment is presented in Section V. Finally, in conclusion, lesson learnt, and future research direction to further enhance the features and capabilities of the perception system is presented in Section VI.

II. Related Works

The integration of sensors and perception algorithms is crucial in AV systems for accurately identifying and categorizing objects like traffic lights, traffic signs, and road markings on public roads. These elements are essential for road safety and traffic management. However, real-time detection of these elements is challenging due to varying environmental conditions, such as changes in lighting, occlusions, image quality issues, motion blur, and glare. These challenges highlight the need for selecting appropriate sensor modalities (e.g., cameras, LiDAR, RADAR) and optimizing perception algorithms to enhance the reliability of AV systems. Several studies have focused on overcoming these difficulties, emphasizing the importance of real-time automation in AVs, particularly at traffic junctions, pedestrian crossings, and roundabouts, where accurate detection and recognition are crucial for preventing collisions and ensuring pedestrian safety.

Early research in TLR relied on rule-based and classical machine learning detectors, which were eventually outperformed by deep learning-based methods. These learning-based detectors significantly improved precision and recall, demonstrating the potential of deep learning for enhancing detection accuracy [15]. The use of stereo cameras has also been proposed to enhance tracking capabilities, particularly for object localization and motion estimation [16]. Addressing data imbalance issues in training datasets is crucial, as seen with the LARA and LISA traffic light databases, where uneven data distribution across traffic light states requires data-centric approaches to improve model performance [17].

Deep learning methods, especially Convolutional Neural Networks (CNNs) and YOLO detectors, have shown promise in TLR. For example, YOLOv3 achieved an AUC of 90.49% and a precision of 50.32% on the LISA dataset [18]. Suggestions for improvement include testing models under nighttime conditions and using ensemble methods like SSD and R-FCN [19]. Advanced methods have been explored, such as using CNNs for traffic light color recognition and integrating Faster R-CNN with k-means clustering, achieving an average precision (AP) of 83%, which increased to 90% for objects larger than 8 pixels [20]. However, challenges remain, including high false positive rates and ensuring consistent classification [21]. The development of large-scale, high-variance datasets like the DriveU Traffic Light (DTLD) dataset, recorded across 11 cities, has been a significant contribution, providing valuable resources for training and evaluating CNN-based models [22].

Traffic road signs provide crucial information that is essential for the decision-making process and safety of autonomous vehicles (AVs). These signs, including speed limits, danger warnings, and directional guidance, are vital for ensuring safe navigation. For instance, when an AV detects a "School ahead" or "Hospital ahead" sign, it can adjust its speed and exercise increased caution. However, detecting traffic signs is challenging due to factors like lighting variations, changes in scale, weather conditions, occlusions, and rotations. Various approaches, including traditional object detectors like Support Vector Machines (SVMs) and pattern matching techniques, have been explored to address these challenges [23]. Despite their application, these methods often struggle with detecting small-scale traffic signs or performing well under difficult conditions [24].

Recent advancements in traffic sign detection have leveraged CNNs and YOLO detectors. While region-based networks and one-stage detectors face limitations in detecting small-scale signs, Region-Proposal Networks (RPNs) have shown superior performance [25]. The integration of Inception V2 for feature extraction has led to competitive results in benchmarks like the German Traffic Sign Detection Benchmark (GTSDB) [26]. A notable method involved generating traffic sign proposals using a color probability model and the Maximally Stable Extremal Region (MSER) detector, followed by an SVM classifier to filter out false positives and categorize signs [27]. Another approach introduced "Capsule Networks," which capture complex spatial relationships, enhancing detection accuracy [28]. Additionally, a model using a single CNN to estimate the location and boundary of traffic signs has improved performance in detecting small and occluded signs [29].

III. Traffic Light and Traffic Road Sign Detection Development for AV Research Platform

The TLR and TSD systems are integral components of the perception system in the AV test vehicle. TLR processes camera or sensor data to detect and interpret traffic light signals, coordinating with the vehicle's path planning and decision-making systems to ensure appropriate actions like stopping or proceeding. TSD captures and processes images to identify and interpret road signs, influencing the vehicle's driving behavior. These systems were tested and validated in the AV research platform's test vehicle at the CETRAN [30] proving test track and along public roads in Cleantech Park and NTU Singapore campus.

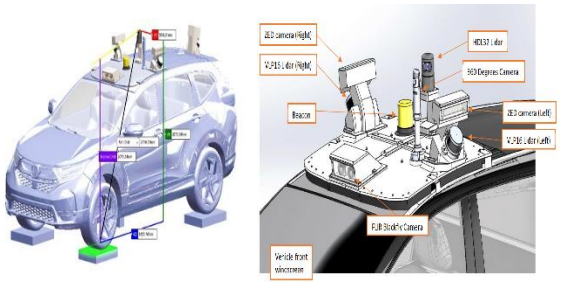
A. AV Research Platform, Test Vehicle & Test Region

The AV research platform uses a Honda CR-V Hybrid Electric Vehicle (HEV) as a medium-size SUV test vehicle to develop and test the AV prototype's sensor and perception systems. The platform integrates high-performance, reliable hardware components with reference autonomous driving software (ADS), ensuring compatibility and robustness. Refer to Fig. 2 for the AV prototype research platform test vehicle.

1) Hardware: The selection process prioritizes commercial off-the-shelf (COTS) items from OEM manufacturers, certified for AV development. It emphasizes CPU and GPU capabilities for efficient parallel processing of sensor data and decision-making tasks. The custom-built industrial PC with an Intel Core i9, 64GB DDR4 RAM with NVIDIA GPU RTX 3080 and AGX Orin is recommended for handling deep learning-based perception algorithms and real-time image processing. The GPU's energy-efficient design and small form factor are well-suited for complex algorithms, while its compatibility with ROS ensures seamless integration into the vehicle.

2) Sensors suite: The sensor perception system integrates various key sensors, including LiDAR, cameras, GNSS+RTK, IMU, and ultrasonic sensors, to enable comprehensive environmental awareness. The LiDAR provides 360º 3D images with high accuracy and long-range sensing, while the GNSS system combined with the local RTK network SiReNT, ensures precise positioning. The IMU offers reliable measurements of angle, angular velocity, and acceleration. Visual perception is achieved through a FLIR Blackfly camera mounted on the front view of the vehicle, providing short- and long-range 2D images, and a Mynteye/ZED stereo camera for full-field 3D measurements. Ultrasonic sensors enhance distance detection, contributing to the AV's robust perception capabilities, particularly in environments with complex traffic lights and road signs. Refer to Table I for the Vision Sensors Filed of View (FoV) device measurements.

3) Software stack: The software stack, built on ROS and running on Ubuntu 18.04, includes sensing, perception, planning, and control software packages for ADS, enabling SAE level 3 autonomy. It processes real-time data from front-facing cameras, LiDAR, and GNSS+IMU+RTK for environmental awareness, employing deep learning algorithms like YOLOv3 for object and traffic light signal recognition. Object tracking uses 2D and 3D data fusion from vision and LiDAR detectors to prevent collisions. The perception system predictions guide the ADS in making decisions regarding objects, obstacles, and traffic signals.

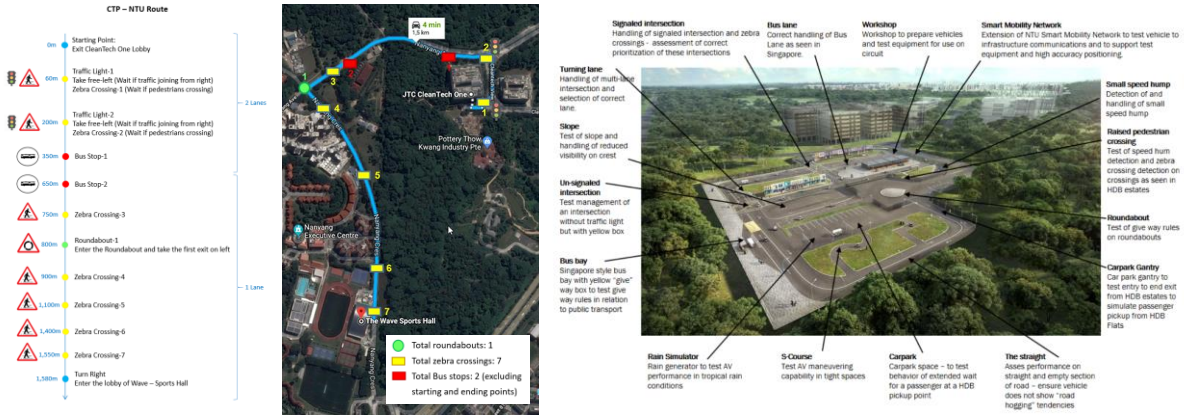


(a) (b)

Fig. 2. (a) The AV research test vehicle equipped with (b) roof-mounted sensor suite for detecting obstacles, pedestrians, traffic lights, road signs, and vulnerable road users. This sensor data supports decision-making, navigation, and control in complex environments.

TABLE I. Field of View (FoV) of Vision Sensors

Sensor	FoV Vertical	FoV Horizontal
LIDAR - HDL 32	+10 to -30 Degrees	360 Degrees
LIDAR - VLP16	+15 to -15 Degrees	360 Degrees
CAMERA - FLIR Blackfly	45 Degrees	60 Degrees
CAMERA - ZED 2	120 Degrees	120 Degrees



(a) (b)

Fig. 3. The image shows designated AV test regions at NTU's campus and CETRAN proving ground. (a) The map highlights key testing locations such as zebra crossings, bus stops, and intersections along the NTU route. (b) The CETRAN facility showcases urban driving scenarios like S-curves, carpark gantries, and smart mobility networks, critical for evaluating AV performance.

4) Testing region. The CETRAN test track facility, managed by NTU at the NTU Smart Campus Cleantech Park in Jurong Innovation District replicates urban road conditions in Singapore. This facility includes the traffic lights and road signs, bus stops, pedestrian crossings, and tropical weather scenarios like heavy rain. This facility allows for controlled, realistic testing of AVs, providing the flexibility to experiment with various AV features without on-road traffic risks. It has a proving ground for assessing AV performance and safety, validating designs for transport, and guiding AV development and certification. The site supports the NTU AV project team's progress toward Level 3 autonomy, with additional trials conducted on selected NTU campus roads approved by the transportation regulator. These trials are part of a milestone testing regime required before public road trials, focusing on mixed bi-directional traffic routes at NTU Clean Tech Park (CTP) and NTU The Wave Sports Centre. The trials routes shown in Fig. 3 covered areas (a) NTU - CTP - NTU The Wave Sports Centre, (b) CETRAN.

B. Traffic Lights Detection and Recognition

The TLR system is vital for the safe navigation of AVs especially in urban environments. It primarily uses camera detection, with sensor data fusion to enhance reliability and reduce false positives. The system applies pre-processing steps like color segmentation and edge detection to images before isolating regions of interest (ROI) to focus on likely by LiDAR and radar for obstacle traffic light locations, reducing computational load inputs, supplemented computational load inputs, supplemented. Initially, YOLOv3 was used for real-time detection and classification of traffic lights, balancing speed, and accuracy. YOLOv3 predicts bounding boxes and classifies traffic light states with confidence scores to minimize false positives, with temporal smoothing algorithms ensuring consistent recognition across frames. The detection process integrates with the path planning module to send a stop or go commands based on traffic light status.

To further enhance detection accuracy and performance, newer versions of YOLOv5 and YOLOv7 were introduced. YOLOv5 improved feature extraction and accuracy for small objects, while YOLOv7 offered enhanced detection capabilities with advanced backbone architectures and layer aggregation, improving performance in complex environments. Both versions provided improved computational efficiency, crucial for real-time operations, and better generalization across diverse conditions, ensuring high reliability. Their modular design also allows for future enhancements, making the system adaptable to evolving AV technology needs.

C. Traffic Road Sign Detection

The TSD system is vital for autonomous driving, enabling vehicles to accurately interpret and respond to road signs, ensuring safe and compliant navigation. The process starts with integrating camera sensors that capture real-time video feeds for detecting road signs. These images undergo pre-processing, such as text and color normalization and edge enhancement, to emphasize features relevant to road signs. The system isolates regions of interest (ROI) to optimize the detection process by focusing computational resources on areas likely to contain road signs.

We run an experiment for YOLOv5 and was chosen for its efficiency in real time object detection, offering a good balance between speed and accuracy. YOLOv5 predicts bounding boxes and classifies road signs into categories like stop signs and speed limits. The recognition module interprets these signs and passes the information to the decision-making module, which adjusts the vehicle's actions accordingly, such as modifying speed based on detected speed limit signs. To further improve detection accuracy and reliability, YOLOv7 was considered, offering enhanced performance and better generalization across different environmental conditions, making the detection system more robust in diverse scenarios. The detection and decision-making process involves sensor integration, pre-processing, detection architecture (starting with YOLOv5 and progressing to YOLOv7), and recognition, with each stage contributing to the system's overall efficiency and performance in AV) applications.

IV. METHODOLOGY

This section outlines the methodology employed for the development and evaluation of TLR and TSD and recognition systems in the AV research platform sensor and perception systems using custom-trained YOLO variant models. The methodology is divided into five key components: Dataset Selection, Preparation, and Preprocessing; Model Implementation; Training and Optimization; Performance Evaluation; and Comparative Analysis; and Experiment.

A. Dataset Selection, Preparation, and Preprocessing

The dataset selection, preparation, and preprocessing involved utilizing the nuances images with 2D annotations [32] and the Singapore Traffic Road Sign Dataset [33]. These datasets provide a comprehensive set of images capturing various Singapore public roads, featuring traffic lights and road signs under different environmental conditions, including daytime, night time, and light rain [34]. This diversity in the dataset is crucial for ensuring the model can generalize across various scenarios, enhancing the robustness of the AV perception system.

The custom and curated datasets used include:

- Traffic Light Dataset: Consists of approximately 700,000 image frames annotated with 17 object labels, including traffic lights with signal color status.
- Traffic Road Sign Dataset: Contains 100,000 image frames with 2,549 labels in 1,778 images across 22 class labels.

1) Data preparation is essential for ensuring effective learning and reliable real-world performance. Annotation Format was converted to YOLO-compatible formats, including class, and bounding box coordinates. Data Splitting involved dividing the datasets into training (75%), validation (15%), and testing (10%) sets, which allowed for the assessment of the model's generalization capabilities on unseen data. Class Balance was analyzed, and techniques like oversampling and class weighting were employed to address imbalances. Strategies included combining similar road sign classes and augmenting underrepresented nighttime images through targeted data augmentation and synthetic data generation. This approach aimed to balance the dataset and enhance the model's generalization ability, resulting in 1,321 training images, 279 validation images, and 178 testing images, which helped reduce the risk of overfitting.

2) Preprocessing steps involved normalization, resizing, and augmentation to further improve the model's generalization capabilities. Normalization scaled pixel values to a common range [0, 1] by dividing them by 255, ensuring consistent input feature scaling, which is crucial for the convergence of gradient-based optimization algorithms. Image Augmentation included transformations such as rotation, scaling, flipping, and color adjustments (brightness, contrast, hue, and saturation) to increase dataset diversity and simulate real-world conditions, including varying lighting and weather scenarios, to improve performance under challenging conditions like low light. A significant challenge during augmentation was the introduction of unrealistic distortions, leading to overfitting, which was mitigated by carefully tuning augmentation parameters and conducting a pilot run to evaluate the effects of each technique on model performance.

B. Model Implementation

The model implementation phase involved transforming the theoretical framework into a functional system by selecting appropriate model architectures, configuring them to meet task-specific requirements, and setting up the training environment.

1) Model architecture customization: The YOLO models—YOLOv3, YOLOv5, and YOLOv7—were selected for their optimal balance between speed and accuracy, making them well-suited for real-time applications. To enhance detection accuracy, predefined anchor boxes were recalculated using k-means clustering to better match the aspect ratios of objects like traffic lights and road signs, resolving issues with mismatched default settings. Adjustments to the number of convolutional layers and filters were made to balance accuracy and inference speed, with overfitting being mitigated by introducing dropout layers and L2 regularization, particularly in YOLOv7. Additionally, modifications to the head and neck structure, including the feature pyramid network (FPN) layers, improved detection performance for objects of varying sizes, such as distant traffic lights. Hyperparameters such as learning rate, batch size, and momentum were fine-tuned, with initial training instability due to a high learning rate being resolved through the use of a cosine annealing learning rate schedule.

2) Pre-training, environmental setup, and model initialization: The models were initialized with pre-trained weights from YOLO models trained on the COCO dataset, which were then fine-tuned on the specific traffic light and road sign dataset. The training environment was set up with GPUs, such as the Nvidia Titan RTX, and included essential software dependencies like PyTorch and CUDA to ensure efficient training. Challenges with GPU memory limitations during large batch size training were addressed using mixed-precision training, which reduced memory usage while maintaining computational efficiency. Proper initialization and checkpointing were also crucial for ensuring stable training throughout the process.

C. Training and Optimization

The goal of the training and optimization process was to maximize model performance, ensuring high accuracy and robustness in detecting and classifying traffic lights and road signs under various conditions. The process involved several key steps:

1) Training process: The training process began with the initialization of pre-trained weights, such as yolov5l.pt, which provided a foundation for fine-tuning the model according to the specific dataset requirements. Data augmentation techniques, including random cropping, scaling, and flipping, were employed to improve the model's generalization across various scenarios. Hyperparameters, such as learning rate,

momentum, batch size, and the number of epochs, were meticulously tuned to strike a balance between training speed and model accuracy. The Stochastic Gradient Descent (SGD) optimizer with momentum was used to update model weights, aiding in faster convergence. The loss function was carefully designed with components for object classification, bounding box regression, and objectness score, with balanced weighting to ensure the model-maintained focus on both localization and classification accuracy. A key challenge in this process was preventing the model from disproportionately prioritizing one aspect over the other, which was effectively managed by adjusting the loss function weights through experimentation.

2) Optimization techniques: The optimization techniques employed focused on enhancing convergence speed, stability, and generalization. A cosine annealing scheduler was used for dynamic learning rate adjustments, incorporating a warm-up phase at the start to stabilize early training and prevent gradient explosions. Regularization methods, such as weight decay and dropout, were implemented to mitigate overfitting, especially during later epochs on smaller datasets. Early stopping and the use of strong dropout layers were particularly effective in preserving model performance. Data augmentation through Mosaic was crucial in increasing dataset diversity, addressing the challenge of limited data for rare road signs. Early stopping and checkpointing strategies helped avoid overfitting and preserved optimal model weights when further training risked degrading validation performance. Mixed precision training on Nvidia GPUs efficiently managed memory resources, reduced training time, and maintained high model performance. Additionally, a grid-based hyperparameter search combined with cross-validation was used to identify the best combination of learning rates, batch sizes, and momentum values, resulting in improved model convergence.

D. Performance Evaluation

The performance evaluation phase focused on assessing the effectiveness and reliability of the trained models using a comprehensive set of evaluation metrics and scenario-based testing.

1) Evaluation metrics such as mAP, IoU, and F1 score were used to evaluate the models. mAP provided an overall measure of detection performance across different object classes by averaging the precision-recall curve. IoU quantified the overlap between predicted bounding boxes and ground truth, giving insight into localization accuracy. The F1 score, which balances precision and recall, was critical for assessing the model's ability to accurately classify and detect objects without missing or falsely detecting them.

2) Scenario-based testing trained models were deployed in various real-world conditions to evaluate their robustness and generalization capabilities. Testing was conducted in diverse environments, including urban settings with varying lighting, weather conditions, and traffic dynamics. The models were tested on the CETRAN proving test track and in multiple regions of the NTU campus, each presenting challenges like mixed bi-directional traffic, complex road layouts, and variable lighting conditions. This approach ensured the models performed reliably across a range of scenarios, reflecting the diverse conditions an AV might encounter in real-world operations.

E. Comparative Analysis

1) Quantitative analysis: The quantitative analysis compared the performance of YOLOv3, YOLOv5, and YOLOv7 using metrics such as mAP, IoU, F1 score, and inference speed (FPS). YOLOv5 and YOLOv7 demonstrated superior mAP scores compared to YOLOv3, particularly excelling in detecting smaller objects and distinguishing between different traffic signal colors and road signs, which are crucial for AV perception systems. YOLOv7 outperformed both YOLOv3 and YOLOv5 in terms of inference speed, making it ideal for real-time applications that require quick decision-making. Additionally, YOLOv5 offered a good balance between speed and accuracy, making it suitable for scenarios where both factors are important. YOLOv7 also exhibited a more optimized trade-off between accuracy and model size, enabling it to run efficiently on AV hardware platforms like the Nvidia Titan RTX and Nuvo-6108GC.

2) Qualitative analysis: The qualitative analysis focused on the models' real-world performance and their ability to generalize to diverse and challenging environments. YOLOv5 and YOLOv7 exhibited better robustness across various environmental conditions, including different lighting and weather scenarios, compared to YOLOv3. YOLOv7 produced fewer false positives and negatives, particularly in cluttered scenes with multiple objects, while YOLOv5 performed well but showed slightly more false negatives in low-light conditions. In scenario-specific performance, YOLOv7 was preferred for scenarios requiring high accuracy, such as detecting smaller, less visible road signs, whereas YOLOv5 provided balanced performance across different test regions like the NTU campus and CETRAN proving test track.

F. Experiment

1) Environmental setup: The environment setup involved both hardware and software configurations to optimize the training and deployment of the models. The models were trained offline using an Nvidia Titan RTX GPU, selected for its ability to manage large datasets and complex models like YOLOv3, YOLOv5, and YOLOv7. For real-time inference and performance evaluation during field testing, the models were deployed on the AV test vehicle's perception system, which is equipped with an Nvidia RTX GPU within the Nuvo-6108GC. The training was conducted using the PyTorch framework, supported by key libraries such as OpenCV, TensorBoard, and YOLO-specific tools for data augmentation and anchor generation. A Linux-based OS, optimized for CUDA operations, was used for both offline and onboard systems. Data management involved storing datasets on high-speed SSDs to reduce loading times during training, and for field testing,

model weights and necessary datasets were preloaded onto the AV's onboard system.

2) Model implementation and training: The model implementation and training process began with the utilization of pre-trained YOLOv3, YOLOv5, and YOLOv7 models, originally trained on the COCO dataset. These models served as a foundation and were fine-tuned for specific tasks related to TLR and TSD. Custom training involved loading datasets consisting of 700,000 image frames for traffic lights and 100,000 for road signs into the training environment. The fine-tuning process focused on optimizing hyperparameters and applying data augmentation techniques like mosaic and mixup to enhance model robustness. Continuous validation monitoring was crucial in detecting overfitting and adjusting training strategies as needed. Optimization techniques included the use of a cosine annealing scheduler for dynamic learning rate adjustments, and regularization techniques like weight decay and dropout to prevent overfitting. Early stopping was employed to avoid unnecessary computation, and mixed precision training was utilized to efficiently manage memory resources, allowing for larger batch sizes without compromising training speed or accuracy.

3) Deployment and field testing: The deployment and field-testing phase involved transferring the optimized model weights to the AV perception system, enabling the AV test vehicle to process and respond to visual inputs in real-time during field tests. Field testing was conducted in key regions within NTU's campus, including the CETRAN proving test track and NTU-Clean Tech Park, where the environments presented complex traffic scenarios to evaluate the models' traffic light and road sign recognition capabilities. During these tests, real-time logging and analysis of model predictions were carried out, focusing on critical metrics such as detection accuracy, false positive/negative rates, and inference speed, to assess the performance of the deployed models.

4) Analysis and iteration: The analysis and iteration phase involved post-field-testing analysis, where performance data from each test was reviewed and verified to identify weaknesses and areas for improvement, especially in cases where the model struggled to correctly identify traffic lights or road signs under varying environmental conditions. Based on these analyses, the models were refined further through fine-tuning, retraining with augmented datasets, or architectural adjustments to better align with AV system requirements. Once optimized and validated, the models were deployed for long-term testing and evaluation in the AV test vehicle, with continuous monitoring and updates as new data and scenarios emerged.

V. Results and Discussion

A. Detection Model for Traffic Lights

The training process for the TLR model demonstrated a progressive improvement across multiple metrics as training epochs progressed. The loss curves, including box loss, object loss, and classification loss, reveal the training dynamics and model convergence behavior. The box loss started at a relatively, higher value and gradually decreased over the training epochs, indicating that the model was improving in accurately predicting the bounding boxes for detected objects. The object loss also followed a declining trend, suggesting that the model was becoming better at identifying whether an object exists in each bounding box. Similarly, the classification loss reduced as training continued, which shows that the model's ability to correctly classify the detected objects improved over time.

1) Epochs of significant improvements at *epoch 30-50*: Notable improvements were observed in this range, where the precision and recall metrics showed significant jumps. The loss curves also depicted steeper declines during these epochs, indicating faster convergence. *Epoch 90-110:* Another period of significant improvement occurred, particularly in the mAP metrics. The model's ability to detect objects with higher accuracy at varying IoU thresholds improved markedly.

2) Challenges encountered toward the later stages of training, particularly after epoch 120, signs of overfitting began to emerge. This was indicated by a plateau in validation metrics such as mAP@0.5 and mAP@0.5:0.95, while the training metrics continued to improve. Overfitting was also suggested by the increase in validation losses despite the continual decrease in training losses. In the early epochs (0-20), the model showed some signs of underfitting, where the precision and recall metrics were relatively low, and the loss values were high. This phase was marked by a slower reduction in losses and a gradual improvement in detection metrics.

The initial training results for the TLR model reveal a detailed analysis of the model's performance in recognizing various traffic light states, including "RedLeft," "Red," "GreenLeft," and "Yellow." The model showed moderate accuracy, with a precision of approximately 86% and a recall of about 80%. The mAP at an IoU threshold of 0.5 (mAP@0.5) reached around 86%, while the more stringent mAP@0.5:0.95 was approximately 65%. These results indicate that while the model performs adequately in many cases, there is room for improvement, particularly in distinguishing between similar traffic light states, which is crucial for making safe and accurate driving decisions in real-world AV applications. Refer to Table II for the summary of the training model results. The confusion matrix shows that the model was relatively accurate in distinguishing between different traffic light states, but there were instances of misclassification, particularly between visually similar states. For example, the model sometimes confused "RedLeft" with "Red" or "GreenRight" with "GreenLeft." The F1-Confidence curve further illustrates the model's precision and recall balance across different confidence thresholds, showing an optimal F1 score of 0.83 a confidence threshold of 0.65, which suggests the model can be further optimized to improve overall performance. Another observation in the initial model tested offline using recorded image bag files, where several issues were identified. This likely due to instance distribution imbalance. Specifically, there was an imbalance in the number of samples for objects and traffic lights, as well as between daytime and night time samples. Although the model was able to detect some of the traffic light status, it failed to identify the green right signal due to its proximity to another

green traffic light signal. Another observation on the system fails to identify a traffic light status that is present in the scene. This can be problematic because it means that the detector missed an object it was supposed to detect. This can add to the false negative detection. Refer to Fig. 4 and Fig. 5 for the observation.

TABLE II. SUMMARY OF THE INITITAL TRAINING RESULTS TRAFFIC LIGHT DETECTION MODEL – YOLOV3

Metric	Results
Dataset	TLR Dataset
Precision	86%
Recall	80%
F1 score	0.83 at 0.65
mAP @0.5	86%
mAP @0.95	65%
Inference Speed (ms/image)	12 ms/frame
Model Size (Parameters)	44M

(a) (b)

Fig. 4. The image compares traffic light detection with ground truth validation. (a) The original scene without detected objects. (b) The system successfully detects and labels two red traffic lights with confidence scores of 0.63 and 0.53. Ground truth annotations are used to validate the system's accuracy in real-world conditions.

(a) (b)

Fig. 5. The image demonstrates the system's detection and recognition capabilities under challenging conditions. (a) Left: the system detects multiple red traffic lights and a 'mandatory left' sign during heavy rain. (b) Right: in low visibility, the system accurately identifies a construction site and split signs but misses a 'give way' sign and a red traffic light (false negatives highlighted in red).

The model's inference speed is an essential factor, particularly for real-time applications like autonomous driving. The model demonstrated an inference speed of approximately 12 milliseconds per image, which is fast enough for real-time detection tasks. Additionally, the model parameter size is around 44 million (44M), striking a balance between complexity and computational efficiency.

Evaluating the performance of a model for deployment in an AV it is important to consider the trade-offs between accuracy, inference speed, and model size. In this case, the model's inference speed of 12 milliseconds per image is sufficient for real-time applications, but the trade-off is a relatively moderate accuracy, particularly under more stringent evaluation conditions (mAP@0.5:0.95). For deployment in an AV test vehicle, a balance must be struck between accuracy and speed. If higher accuracy is required, more complex models or ensemble methods could be used, though this would likely reduce inference speed. Conversely, if speed is prioritized, a more streamlined model might be necessary, though this could compromise accuracy.

The TLR model exhibits moderate accuracy and a fast inference speed, making it suitable for real-time applications in AV. However, the performance of the model could be improved by addressing class imbalances, enhancing data augmentation techniques to better handle variations in lighting and angles, and fine-tuning the model's hyperparameters. Additionally, considering more advanced architectures or employing ensemble learning methods could help improve accuracy and robustness, particularly in distinguishing between visually similar traffic light states. For deployment in an AV test vehicle, this model could serve as a baseline, but further refinements may be necessary depending on the specific requirements for accuracy and speed in the target application. Balancing the trade-offs between accuracy and inference speed will be critical to ensuring that the model meets the operational demands of real-time autonomous driving. Refer to Table III for the summary of the TLR performance results.

B. Detection Model for Traffic Road Signs

The initial training results for the three YOLOv5 models—YOLOv5s, YOLOv5m, and YOLOv5l—demonstrate different strengths in detecting and recognizing traffic signs, varying in accuracy, speed, and computational demands. As training progresses, there is a consistent decrease in losses related to bounding box regression, object classification, and class prediction, indicating that the models are effectively learning. The reduction in metrics like train/box loss, which starts at 0.11157 and decreases to around 0.0235, suggests improved accuracy in predicting object locations and class labels.

Performance metrics such as precision, recall, and mean Average Precision (mAP) also improve throughout training. Precision stabilizes and exceeds 0.93, while recall shows a steady increase, indicating the model's ability to capture more relevant objects. The mAP metrics, including mAP_0.5, reach approximately 0.96 by the end of training, reflecting the model's effectiveness across various IoU thresholds. Refer to the summary Table IV training and Table V performance.

TABLE III. TRAFFIC LIGHT DETECTION PERFORMANCE RESULTS

Model	Precision (%)	Recall (%)	F1 Score (%)	mAP@0.5	mAP@0.95	Inference Speed (fps)	Model Size
YOLOv3	86	80	83	0.86	0.65	12	240
YOLOv5	90	87	88.5	0.90	0.68	28	140
YOLOv7	93	90	91	0.93	0.70	22	120

TABLE IV. SUMMARY OF TRAINING RESULTS TRAFFIC ROAD SIGN – YOLOV5

Model	Batch Size	Precision	Recall	mAP @0.5	mAP @0.95	Inference Speed (ms/image)	Model Size (Parameters)
YOLOv5s	16	90%	85%	90%	58%	7 ms	7M
YOLOv5m	32	92%	88%	93%	61%	10 ms	21M
YOLOv5l	64	93%	87%	94%	60%	14 ms	47M

TABLE V. TRAFFIC ROAD SIGN DETECTION PERFORMANCE RESULTS

Model	Precision (%)	Recall (%)	F1 Score (%)	mAP@0.5	mAP@0.95	Inference Speed (fps)	Model Size
YOLOv3	92	90	91	0.92	0.70	25	240
YOLOv5	94	92	93	0.94	0.72	28	140
YOLOv7	96	94	95	0.96	0.74	22	120

Validation losses, which decrease consistently alongside training losses, indicate strong generalization to unseen data and a low risk of overfitting. The learning rates, initially set high and gradually reduced, facilitate faster convergence early on and fine-tuning later in the training process.

Despite the strong performance, further improvements could be achieved through advanced augmentation techniques, fine-tuning hyperparameters, or applying model pruning and quantization to reduce model size and inference time. These adjustments would enhance the model's suitability for deployment in resource-constrained environments like an AV test vehicle, where the trade-off between accuracy and inference speed is crucial.

C. Field Testing in Real-World Scenarios

Field testing of TLR and road TSD models was conducted on the AV test vehicle using the Nuvo-6108GC PC environment equipped with an NVIDIA RTX GPU. The goal was to evaluate the model's performance in real-world scenarios reflective of Singapore's urban conditions. Testing occurred across test regions: the CETRAN proving test track, JTC Cleantech Park, and NTU's campus, each presenting distinct challenges. The CETRAN test track, designed for AV testing, replicates various Singaporean urban road elements, offering a controlled environment for rigorous testing without the risks of live traffic. This site allowed for comprehensive assessment of the AV's manoeuvrability through city driving scenarios [31].

1) Test regions and specific challenges. CETRAN facility provided a geofenced, controlled environment showcasing urban scenarios like S-curves and carpark gantries, crucial for evaluating AV performance. Traffic lights and road signs were readily available for evaluation without traffic interference. NTU-Clean Tech Park (CTP) - The Wave Sports and Recreation Centre featured mixed bi-directional traffic, creating a complex environment to test the model's ability to accurately detect and classify traffic lights and road signs amidst dynamic vehicle movements and varying traffic density. The Nuvo-6108GC, with its NVIDIA RTX GPU, supported the real-time processing needs of advanced deep learning models, achieving an average inference speed of 12ms per frame, suitable for urban traffic scenarios. The model demonstrated robust detection, maintaining high precision and recall metrics across various conditions, including challenging lighting and complex traffic scenarios.

2) Challenges observed during field testing. The model performed well overall, but low-light scenarios occasionally reduced detection accuracy for certain road signs. The model navigated intersections and recognized pedestrian crossings effectively, though closely spaced traffic lights sometimes caused minor detection delays. The model maintained detection capabilities in light rain, but heavy rain introduced reflections that occasionally confused the perception system.

D. Lessons Learned

The performance evaluation of the model development and deployment in the AV research platform focused on enhancing safety and robustness, particularly in image recognition, motion speed profiles, and obstacle detection at critical areas like junctions and pedestrian crossings. Several test scenarios, including encounters with pedestrians and other vehicles, provided insights that guided improvements in design and safety approaches.

1) Model development and training revealed that the choice of model architecture significantly impacted the balance between detection accuracy and inference speed. YOLOv7 excelled in complex urban environments but required higher computational resources. Data augmentation techniques like mosaic and mixup were essential in improving the model's robustness across diverse conditions, though issues with data

imbalance, particularly in night time scenarios, highlighted the need for a balanced training dataset.

2) Deployment and real-time inference showed that fine-tuning was necessary to optimize real-time performance. While YOLOv7 offered the highest accuracy, its computational demands resulted in slightly slower inference times, particularly in high-traffic environments where quick decision-making was crucial. This trade-off emphasized the need for further optimization to handle scenarios with high visual complexity.

3) Field testing observations across the CETRAN proving test track, NTU-Clean Tech Park, and NTU campus highlighted YOLOv7's superior detection accuracy, especially in identifying over 95% of traffic lights and road signs. However, the model's inference speed sometimes lagged in complex traffic scenarios, making YOLOv5 a more balanced choice in environments requiring rapid processing. False positives and negatives were noted, particularly in varying lighting conditions, underscoring the need for improved robustness against environmental noise and challenging conditions like heavy rain and glare. Our team also tested to other testing site to verify the performance of the object detection using other use case in the golf range [35] and integrated wildlife recreation area [36] how the model performs in a different field testing and observation.

4) Post-field-testing analysis involved a detailed review of model failures, particularly under low-light and adverse weather conditions, which led to higher false-negative rates. The analysis stressed the importance of enhancing data augmentation and possibly integrating additional sensors like LiDAR. While YOLOv7 achieved impressive metrics under ideal conditions (precision of 95%, recall of 93%, F1 score of 94%, and mAP@50 of 0.95), these dropped under challenging conditions, indicating a need for further refinement. Inference speed was 22 fps for YOLOv7, compared to 28 fps for YOLOv5, highlighting the trade-offs between speed and accuracy.

VI. Conclusion

This research made significant advancements in AV perception systems by evaluating performance, comparing YOLO variants, conducting real-world testing, and developing a comprehensive testing framework. YOLOv7 emerged as the best-performing model, achieving a mAP@0.5 of 93% for Traffic Light Recognition (TLR) and 96% for Traffic Sign Detection (TSD), even in challenging environments such as low-light and occlusion scenarios. Its superior precision and recall, with F1 scores of 91% for TLR and 95% for TSD, demonstrated its suitability for real-time AV applications. YOLOv5, while slightly less accurate, provided a strong balance between speed and accuracy, making it adaptable for various conditions.

Real-world testing in urban environments, including the CETRAN proving test track and NTU campus, validated YOLOv7's reliable performance and its readiness for deployment in AV systems. A significant contribution of the study was the development of a scenario-based testing framework that included continuous performance monitoring and model refinement. This framework helped identify and correct model weaknesses, enhancing performance and robustness.

The research offers critical insights for AV developers and researchers, particularly on balancing speed, accuracy, and robustness in real-world applications. Future work could focus on exploring hybrid models that integrate YOLO with Transformer-based architectures, testing in diverse environments such as rural areas and highways, incorporating multi-modal data (LiDAR, RADAR), and exploring adversarial robustness testing.

Acknowledgment

This research acknowledges the AV research team of Energy Research Institute (ERI@N) Nanyang Technological University Singapore.

References

[1] Singapore Police Force, "Annual Road Traffic Situation 2021," [Online].Available:https://www.police.gov.sg/media/D4435F72157942D3B323EE4A507D4CFB.ashx. [Accessed: 11-Aug-2024].

[2] D. Lee, "The Big Read: Amid rising number of traffic accidents, a road culture that's proving hard to change," Channel NewsAsia, 23-Aug-2023.[Online].Available:https://www.channelnewsasia.com/singapore/big-read-rising-traffic-accidents-road-culture-4328841.[Accessed:11-Aug-2024].

[3] World Health Organization, "Global status report on road safety 2023," 2023.[Online].Available:https://www.who.int/publications/i/item/9789240045747. [Accessed: 11-Aug-2024].

[4] M. B. Jensen, M. P. Philipsen, A. Møgelmose, T. B. Moeslund and M. M. Trivedi, "Vision for Looking at Traffic Lights: Issues, Survey, and Perspectives," in IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 7, pp. 1800-1815, July 2016, doi: 10.1109/TITS.2015.2509509.

[5] R. Zhang, K. Zheng, K., P. Shi, Y. Mei, H. Li, T. Qiu, "Traffic Sign Detection Based on the Improved YOLOv5," Appl. Sci. 2023, 13, 9748. https://doi.org/10.3390/app13179748.

[6] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.

[7] W. Liu, et al. , "SSD: Single Shot MultiBox Detector" In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science(), vol 9905. Springer, Cham. https://doi.org/10.1007/978-3-319-46448-0_2.

[8] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91

[9] E. Marti, M. A. de Miguel, F. Garcia and J. Perez, "A Review of Sensor Technologies for Perception in Automated Driving," in IEEE Intelligent Transportation Systems Magazine, vol. 11, no. 4, pp. 94-108, winter 2019, doi: 10.1109/MITS.2019.2907630

[10] H. Zhao et al., "Driving Scenario Perception-Aware Computing System Design in Autonomous Vehicles," 2020 IEEE 38th International Conference on Computer Design (ICCD), Hartford, CT, USA, 2020, pp. 88-95, doi: 10.1109/ICCD50377.2020.00031.

[11] M. J. Shafiee, A. Jeddi, A. Nazemi, P. Fieguth and A. Wong, "Deep Neural Network Perception Models and Robust Autonomous Driving Systems: Practical Solutions for Mitigation and Improvement," in IEEE Signal Processing Magazine, vol. 38, no. 1, pp. 22-30, Jan. 2021, doi: 10.1109/MSP.2020.2982820.

[12] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018. [Online]. Available: https://arxiv.org/abs/1804.02767

[13] G. Jocher, J. Nishimura, and L. Tao, "YOLOv5," Ultralytics, 2020. [Online]. Available: https://docs.ultralytics.com/yolov5/. [Accessed: Jan. 20, 2024

[14] C. -Y. Wang, A. Bochkovskiy and H. -Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 7464-7475, doi: 10.1109/CVPR52729.2023.00721.

[15] J. G. Wang and L. B. Zhou, "Traffic Light Recognition with High Dynamic Range Imaging and Deep Learning," IEEE Trans. Intell. Transp. Syst., vol. 20, no. 4, pp. 1341–1352, Apr. 2019. DOI: 10.1109/TITS.2018.2859226.

[16] M. Bach, D. Stumper, and K. Dietmayer, "Deep Convolutional Traffic Light Recognition for Automated Driving," in IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 2018, vol. 2018-November, pp. 851–858. DOI: 10.1109/ITSC.2018.8569907.

[17] Z. Shi, Z. Zou, and C. Zhang, "Real-time traffic light detection with adaptive background suppression filter," IEEE Trans. Intell. Transp. Syst., vol. 17, no. 3, pp. 690–700, Mar. 2016. DOI: 10.1109/TITS.2015.2480700.

[18] J. Levinson, J. Askeland, J. Dolson, and S. Thrun, "Traffic light mapping, localization, and state detection for autonomous vehicles," in Proceedings - IEEE International Conference on Robotics and Automation, 2011, pp. 5784–5791. DOI: 10.1109/ICRA.2011.5980138.

[19] M. Guo, X. Zhang, D. Li, T. Zhang, and L. V. An, "Traffic light detection and recognition for autonomous vehicles," The Journal of China Universities of Posts and Telecommunications, vol. 22, no. 1, pp. 50–56, Feb. 2015. DOI: 10.1016/s1005-8885(15)60624-0.

[20] S. Saini, S. Nikhil, K. R. Konda, H. S. Bharadwaj and N. Ganeshan, "An efficient vision-based traffic light detection and state recognition for autonomous vehicles," 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 2017, pp. 606-611. DOI: 10.1109/IVS.2017.7995785.

[21] Y. Lu, J. Lu, S.-H. Zhang, and P. M. Hall, "Traffic signal detection and classification in street views using an attention model," Computational Visual Media, vol. 4, no. 3, pp. 253–266, Aug. 2018. DOI: 10.1007/s41095-018-0116-x.

[22] M. P. Philipsen, M. B. Jensen, A. Møgelmose, T. B. Moeslund and M. M. Trivedi, "Traffic Light Detection: A Learning Algorithm and Evaluations on Challenging Dataset," 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Gran Canaria, Spain, 2015, pp. 2341-2345. DOI: 10.1109/ITSC.2015.378.

[23] A. Fregin, J. Muller, U. Krebel and K. Dietmayer, "The DriveU Traffic Light Dataset: Introduction and Comparison with Existing Datasets," 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 2018, pp. 3376-3383. DOI: 10.1109/ICRA.2018.8460737.

[24] M. Weber, P. Wolf, and J. M. Zollner, "DeepTLR: A single deep convolutional network for detection and classification of traffic lights," in IEEE Intelligent Vehicles Symposium, Proceedings, 2016, vol. 2016-August, pp. 342–348. DOI: 10.1109/IVS.2016.7535428.

[25] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu, "Traffic-Sign Detection and Classification in the Wild," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, vol. 2016-December, pp. 2110–2118. DOI: 10.1109/CVPR.2016.231.

[26] I. M. Creusen, R. G. J. Wijnhoven, E. Herbschleb, and P. H. N. De With, "Color exploitation in HOG-based traffic sign detection," in Proceedings - International Conference on Image Processing, ICIP, 2010, pp. 2669–2672. DOI: 10.1109/ICIP.2010.5653893.

[27] G. Wang, G. Ren, Z. Wu, Y. Zhao, and L. Jiang, "A robust, coarse-to-fine traffic sign detection method," in Proceedings of the International Joint Conference on Neural Networks, 2013. DOI: 10.1109/IJCNN.2013.6706844.

[28] H. Luo, Y. Yang, B. Tong, F. Wu, and B. Fan, "Traffic Sign Recognition Using a Multi-Task Convolutional Neural Network," IEEE Trans. Intell. Transp. Syst., vol. 19, no. 4, pp. 1100–1111, Apr. 2018. DOI: 10.1109/TITS.2017.2784095.

[29] H. H. Aghdam, E. J. Heravi, and D. Puig, "A practical approach for detection and classification of traffic signs using Convolutional Neural Networks," Robotics and Autonomous Systems, vol. 84, pp. 97–112, Oct. 2016. DOI: 10.1016/j.robot.2016.07.003.

[30] CETRAN, "Centre of Excellence for Testing & Research of Autonomous Vehicles," CETRAN. [Online]. Available: https://cetran.sg/. [Accessed: Jan. 20, 2024].

[31] SAE International, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," SAE Standard J3016_202104, Apr. 2021. DOI: 10.4271/J3016_202104.

[32] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. Ionasescu, A. Y. Zhou, J. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuScenes: A multimodal dataset for autonomous driving," in Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 11618-11628. DOI: 10.1109/CVPR42600.2020.01163.

[33] E. Yan, "Classification and Detection of Singapore Road Traffic Signs,"GitHubRepository,https://github.com/eugeneyan84/Classification-and-Detection-of-Singapore-Road-Traffic-Signs. [Accessed: Jan. 20, 2024].

[34] Y. Yang, H. Luo, H. Xu, and F. Wu, "Towards Real-Time Traffic Sign Detecton and Classification," in IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 7, pp. 2022-2031, July 2016.

[35] J. De Guia et al., "Autonomous Golf Ball Picker: The First in Singapore Golf Environment," 2023 IEEE 12th Global Conference on Consumer Electronics (GCCE), Nara, Japan, 2023, pp. 900-901, doi: 10.1109/GCCE59613.2023.10315528.

[36] J. De Guia et al., "Advancing Safety and Robustness: Perception-Planning System of an Autonomous Vehicle Last Mile Delivery," 2024 IEEE Conference on Artificial Intelligence (CAI), Singapore, Singapore, 2024, pp. 113-118, doi: 10.1109/CAI59869.2024.00260.

Smart X-Ray Geiger Data Logger: An Integrated System for Detection, Control, and Dose Evaluation

Lhoucine Ben Youssef[1], Abdelmajid Bybi[2], Hilal Drissi[3], El Ayachi Chater[4]
Mohammed V University in Rabat, Higher School of Technology of Salé, LASTIMI, Salé, Morocco[1, 3, 4]
Mohammed V University in Rabat, Higher School of Technology of Salé, MEAT, Salé, Morocco[2]

***Abstract*—X-ray dosimetry practices are guided by international standards and regulatory agencies to ensure the safety of patients, radiation workers, and the general public. This paper introduces the Smart X-ray Geiger Data Logger, a comprehensive system designed to enhance radiation safety through integrated detection, control, and dose evaluation. This study is based on the M4011 Geiger-Müller tube, exploiting ionization effects to measure radiation doses accurately. The system features an advanced algorithm for real-time exposure risk assessment, ensuring adherence to safety limits during medical procedures. Equipped with Wi-Fi connectivity, the device facilitates seamless data transmission and integration with centralized databases for comprehensive exposure monitoring and historical data analysis. The MQTT protocol is utilized for secure and efficient data transmission, ensuring the protection of sensitive information. A user-friendly interface provides instant feedback on radiation levels, cumulative doses, and procedural safety, supported by visual indicators and auditory alarms for immediate alerts. Experimental validation demonstrates the system's reliability in various settings, confirming its utility in optimizing radiation protection strategies and fostering safer environments in the healthcare field.**

***Keywords*—X-rays; radiation dose; radiation safety; exposure risk assessment; Geiger-Müller tube; medical imaging; real-time monitoring; smart devices**

I. Introduction

X-rays are a form of electromagnetic radiation with wavelengths shorter than visible light. They were discovered by Wilhelm Conrad Roentgen in 1895. X-rays have the ability to penetrate most substances, and they are widely used in medical imaging, industry, and scientific research. In medical applications, X-rays are commonly used for imaging bones and internal structures, helping diagnose and monitor various medical conditions [1-3].

X-ray dosimetry involves the measurement and assessment of the absorbed dose of X-ray radiation received by human tissue. The primary goal is to quantify the amount of energy deposited in a specific area, which is crucial for various applications, including medical imaging and radiation therapy [4].

Ongoing research aims to develop new dosimetry techniques, materials, and technologies to enhance accuracy, sensitivity, and efficiency in measuring X-ray doses. To illustrate, Fum et al. [5] worked on Monte Carlo (MC) methods in fluoroscopy-guided interventions for clinical dosimetry, with potential advancements in dynamic, deformable phantoms and automated image-based alignment for enhanced patient specificity. To efficiently and expeditiously assess X, γ, and neutron radiation fields concurrently, Yang et al. [6], developed a portable multifunction radiation detection system employing LaBr3(Ce) crystal, LASO neutron detector, and high-range GM counter, data measured transmitted to a PC through a USB interface. Johnson et al. [7] survey cutting-edge approaches for X-ray dose detection, such as semiconductor detectors and digital dosimeters, underscoring their enhanced precision and quicker response times compared to conventional techniques. Furthermore, Garcia et al. [8] assess the implementation of automated X-ray dose tracking systems in hospital settings, revealing a 25% reduction in incidents of excessive radiation exposure among patients. Brown et al. [9] explores advanced dosimetry methods like real-time dosimeters and patient dose mapping, which significantly bolster X-ray imaging safety, resulting in a 40% enhancement in radiation dose control. Moreover, Thomas et al. [10] spotlight recent advancements in X-ray dose monitoring technologies, including wearable dosimeters and machine learning algorithms, which collectively lowered radiation exposure by 20% in clinical environments.

Excessive exposure to X-rays poses significant health risks, including radiation burns, radiation sickness, and an increased risk of cancer. As X-ray technology is widely used in medical imaging, it is crucial to monitor and control the radiation dose received by individuals to ensure their safety. Traditional methods of measuring X-ray exposure can be inadequate or cumbersome, leading to the necessity for a more efficient, accurate, and user-friendly solution. In this context, Wilson et al. [11] investigate the development of portable X-ray dose detectors utilizing microelectromechanical systems (MEMS), demonstrating their high accuracy and ease of use in clinical trials. On the other hand, Smith et al. [12] delve into recent advancements in real-time monitoring of X-ray doses during radiological procedures, utilizing advanced sensor arrays to achieve a 30% decrease in patient radiation exposure. Meanwhile, Robinson et al. [13] analyze the complexities of real-time X-ray dose management through case studies, proposing solutions that integrate dose monitoring systems to enhance safety compliance by 50%. Martinez et al. [14] explore the integration of artificial intelligence to enhance X-ray dose detection and monitoring, revealing that AI algorithms enhance detection accuracy by 25% and reduce false positives by 30%. Lastly, Silva et al. [15] describe the development of low-cost X-ray equipment including an X-ray machine, a Geiger detector, and a goniometer, aimed at facilitating accessible study and experimentation in X-ray physics.

To address the critical need for real-time, precise monitoring of X-ray exposure, we developed the Smart X-ray Geiger Data Logger. This device can provide immediate feedback on radiation doses including instant and cumulative dose, enabling users to take timely actions to minimize exposure. Otherwise, our solution is particularly valuable in medical environments, ensuring that patients and healthcare workers are not subjected to unnecessary radiation. Additionally, the Smart X-ray Geiger Data Logger is a handheld device used to detect, control and evaluate X-ray doses. It operates based on the ionization effect produced when radiation interacts with a gas inside the Geiger-Muller tube. By employing advanced algorithm for exposure risk assessment, the Smart X-ray Geiger Data Logger can improve radiation safety management, protecting health and enhancing overall safety standards.

This paper is structured into two main sections: The first section comprehensively explores the materials and methods employed in the study, detailing the experimental setup, equipment specifications, and data collection procedures. The second section is dedicated to presenting and discussing the results obtained from the experiments. It analyzes the findings in the context of existing solutions, discusses implications for the field, and explores potential avenues for further research.

II. Materials and Methods

A. X-Ray Doses

Dosimeters are employed in medical field to monitor personnel radiation exposure during diagnostic X-ray procedures and in occupational application to ensure that radiation workers are not exposed to harmful levels of radiation. To evaluate the risk of excessive dose absorption, we based on the effective dose of the body, which is a measure used in radiology to estimate the overall risk of harm from ionizing radiation exposure. It takes into account the different sensitivities of various tissues and organs in the body to radiation.

The calculation of effective dose involves three main steps: first, determining organ and tissue weighting factors (Table I), which reflect the varying sensitivities of different tissues and organs to radiation and are set by organizations like the International Commission on Radiological Protection [16]. Second, calculating the absorbed dose, the energy deposited per unit mass in a specific organ or tissue, measured in gray, based on radiation type, energy, and exposure conditions. Finally, the effective dose is obtained by multiplying the absorbed dose in each organ or tissue by its respective weighting factor and summing these values across all exposed organs and tissues.

It's important to note that the effective dose is an estimate of the overall risk and not a direct measure of the biological effect on an individual. It is a useful concept for comparing and managing radiation risks in different exposure scenarios. Absorbed dose is the amount of energy imparted by ionizing radiation per unit mass of irradiated material. It is measured in gray (Gy) in the International System of Units (SI). Absorbed dose can be calculated using dosimeters, devices specifically designed to measure the energy deposited by X-rays in a given material. However, there are various types of dosimeters used in X-ray dosimetry, including ionization chambers, thermoluminescent dosimeters (TLDs), semiconductor detectors, and film badges.

TABLE I. Organ and Tissues Weighting Factor

Tissue or Organ	ICRP 103
Gonads	0.08
Lung	0.12
Colon	0.12
Stomach	0.12
Breast	0.12
Bladder	0.04
Liver	0.04
Esophagus	0.04
Thyroid	0.04
Skin	0.01
Bone surface	0.01
Brain	0.01

The principal function of an X-ray dosimeter based on ionization chambers is to measure and quantify the ionizing radiation exposure from X-rays. Ionization chambers operate by detecting the electrical charge produced when X-ray photons ionize the gas molecules within the chamber [17-19]. Likewise, TLD dosimeters play a crucial role in assessing radiation exposure by utilizing the thermoluminescent properties of certain materials to measure the amount of energy absorbed from ionizing radiation [20-25]. In other hand, Semiconductor detectors for X-rays serve as key tools for detecting and measuring ionizing radiation in various applications [26-29], their high sensitivity, spatial resolution, and fast response time make them valuable in diverse applications, including medical imaging, industrial inspection, and radiation monitoring. Finally, X-ray dosimeters based on film badges are practical tools for monitoring ionizing radiation exposure, offering wearable, cost-effective, and dose-history recording solutions for individuals working in environments where X-rays are present [30-32].

In our case, we used the ionization chamber M4011 Geiger-Müller Tube shown in Fig. 1. The interaction of X-ray photons with gas, leading to the formation of ion pairs (positive and negative charges) within the chamber. An electric field applied within the ionization chamber separates these ion pairs, and the resulting electric current or charge is quantified. This current's magnitude is directly proportional to the degree of ionization, which correlates with the radiation dose. Converting the measured charge into a radiation dose, typically expressed in Sieverts (Sv), enables a quantitative evaluation of absorbed radiation. Renowned for their precision and reliability across a broad spectrum of energies, ionization chambers find application in diverse fields such as medicine, industry, and research.

Fig. 1. M4011 Geiger-Müller tube.

B. X-Ray Doses Detection and Assessment

Dose quantities, such as entrance surface dose (ESD) and organ or effective dose, are often used to assess and report radiation exposure in medical imaging. To effectively measure,

display, control, and monitor these doses, we have developed the Smart X-ray Geiger Data Logger, as shown in Fig. 2.

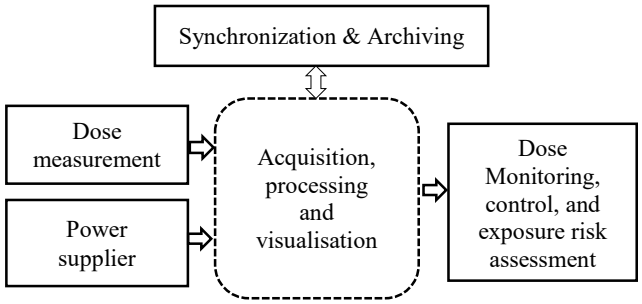


Fig. 2. Smart X-ray Geiger Data Logger for measurement and monitoring of X-ray doses in medical imaging.

The Smart X-ray Geiger Data Logger is an innovative device designed for the collection, control, monitoring, and evaluation of X-ray doses. Its primary purpose is to ensure the safety and protection of individuals from excessive X-ray exposure, thereby reducing the risk of radiation-induced diseases such as cancer.

The device is equipped with its own battery and power adapter, ensuring continuous operation even in the absence of an external power source. It includes memory storage for dose data, allowing offline access to historical exposure information. Built-in Wi-Fi capability allows the device to connect to a database for data acquisition and retrieval. The MQTT protocol is utilized for secure and efficient data transmission, ensuring the protection of sensitive information. An auditory alarm and LED indicators are activated when the absorbed dose exceeds predefined safety limits, providing immediate alerts to the user. Additionally, the device features a specialized algorithm for exposure risk assessment. Before proceeding with a new examination, the algorithm evaluates the risk based on the reference dose for the upcoming exam and the cumulative dose previously received. If the total dose does not exceed the safety limit, the device grants permission for the new exam. Otherwise, it warns the user of the potential risks and suggests alternative procedures.

The Smart X-ray Geiger Data Logger is a versatile tool that empowers individuals to continuously monitor and control their X-ray exposure, ensuring their safety. In medical settings, it helps radiology departments and doctors in making informed decisions about X-ray examinations, enhancing patient safety. Biomedical technicians and engineers benefit from using the device during the installation, maintenance, checkup, startup, and calibration of X-ray equipment, as it helps control their exposure to scattered radiation. Radiology technicians also find it invaluable for managing their radiation exposure during daily operations. The primary objectives of the Smart X-ray Geiger Data Logger are to validate the necessity of X-ray doses, avoid unnecessary exposure, safeguard individuals from excessive radiation, prevent radiation-related diseases, and assist healthcare professionals in optimizing patient care.

In the pursuit of measuring values, our preference leaned towards the M4011 Geiger- Müller tube, chosen for its specific attributes (380-450V, 25 pulse/minutes) [33]. This selection aligns harmoniously with the requirements of our solution. As for the power supply, we integrated a lithium battery featuring 3300 mAh and 3.7V.

Fig. 3 illustrates the diagram of our device and the connections between its various components. At the heart of our system lies the ESP32 board, assuming a central role in managing data acquisition, executing algorithmic processing, and presenting results on the display. Beyond these functions, it acts as a crucial facilitator for seamless communication among the diverse elements within our system. Boasting a robust 32-bit dual-core microcontroller that operates at clock frequencies of up to 240 MHz, the ESP32 board distinguishes itself with its integrated Wi-Fi and Bluetooth capabilities, making it highly suitable for a wide array of IoT (Internet of Things) applications. These boards exhibit varying amounts of Flash memory for program storage and RAM for data storage, the specific sizes of which depend on the model. Noteworthy is the ESP32's intentional design for power efficiency, supporting a range of power modes to optimize energy consumption, particularly beneficial for battery-operated devices [34].

The RadiationD V1.1 board is engineered for precise radiation detection and measurement, incorporating vital features such as powering and adapting the M4011 Geiger-Müller tube. The TP4056 module enhances its functionality with efficient charging of single-cell lithium-ion or lithium-polymer batteries, integrating thermal regulation, overcurrent protection, and reverse polarity protection for safe operation. In Smart X-ray Geiger Datta Logger, the DS3231 real-time clock module ensures precise time synchronization and event logging, facilitating accurate timestamping of radiation readings. This capability is crucial for monitoring radiation dose accumulation over time and correlating exposure data with specific events or patient movements. Additionally, the ILI9341 TFT LCD controller on the board supports essential functions for display management and high-quality graphics rendering, enhancing usability in radiation monitoring applications.

The user interface consists of two main parts. The first part is dedicated to the dashboard, which displays real-time dose measurements, cumulative dose, and the set of exams for the patient. The second part is dedicated to X-ray dose risk assessment. Before any new examination, the patient can select the new exam type, and the system will check the dose associated with this procedure along with the cumulative dose. If the combined dose exceeds the safety limit, the system will alert the user that the selected procedure is risky and suggest choosing an alternative.

The exposure risk assessment is based on X-ray dose references (Table II) for common procedures [35-36]. This algorithm applies the three fundamental principles of radiation protection: Justification, Optimization and Limitation.

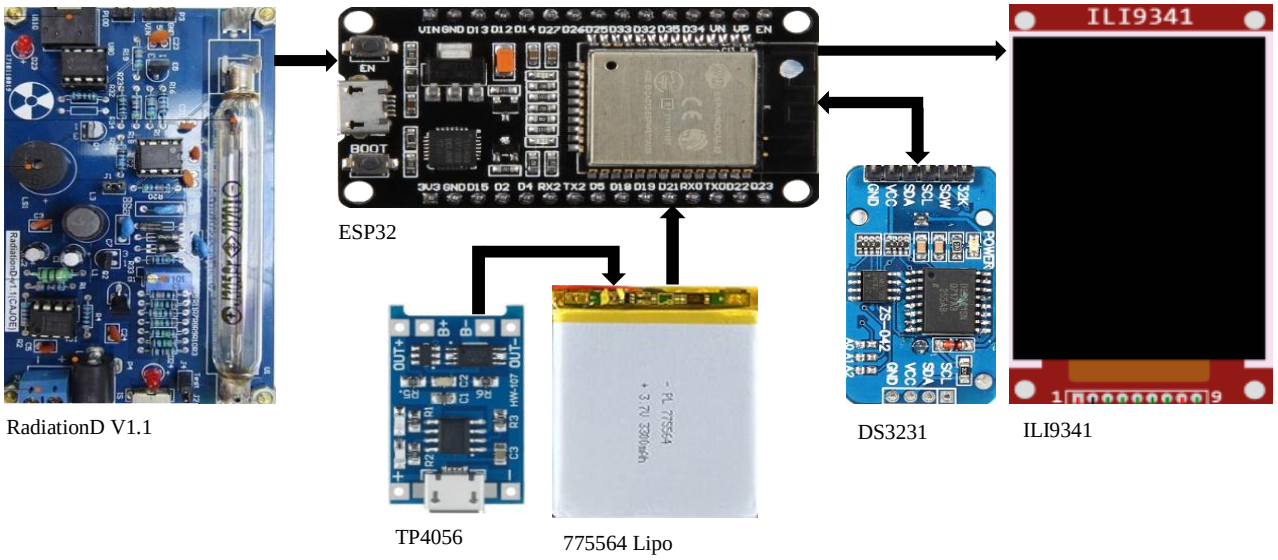


Fig. 3. Smart X-ray Geiger Data Logger diagram.

TABLE II. X-RAY DOSE REFERENCES FOR COMMON PROCEDURES

Procedure	Effective dose mSv
Chest CT	6.1
Abdomen CT	5.3
Brain CT	1.6
Cardiac CT	1.7
Lumbar RS	1.4
Extremity RS	0.1
Chest RS	0.1
Mammography	0.21
Dental	0.25

The Smart X-ray Geiger Data Logger, illustrated in Fig. 4, is engineered to fulfill the objective of measuring, displaying, controlling, and monitoring radiation doses. By embodying these capabilities, it not only upholds stringent principles of radiation safety but also enhances overall radiation protection protocols.

Fig. 5 depicts the flow chart of our solution algorithm. The process begins with powering ON the device and establishing a connection to the database. The user interface then displays real-time dose measurements, cumulative dose, and scheduled exams. The patient selects the type of new exam to be performed, and the device retrieves the dose information associated with the selected exam. The system calculates the total dose by adding the new exam dose to the cumulative dose and evaluates whether this total dose exceeds the safety limit. If it does, a warning message is displayed; if not, the system grants permission for the exam. After the exam, the cumulative dose is updated in the database. The system then checks the updated cumulative dose against the safety limit. If the cumulative dose exceeds the limit, an audible notification is activated and a red LED is turned on. If it does not exceed the limit, the device returns to the dashboard, ready for the next operation, and re-authenticates with the database.

Fig. 4. Smart X-ray Geiger Data Logger for radiation dose measurement and monitoring.

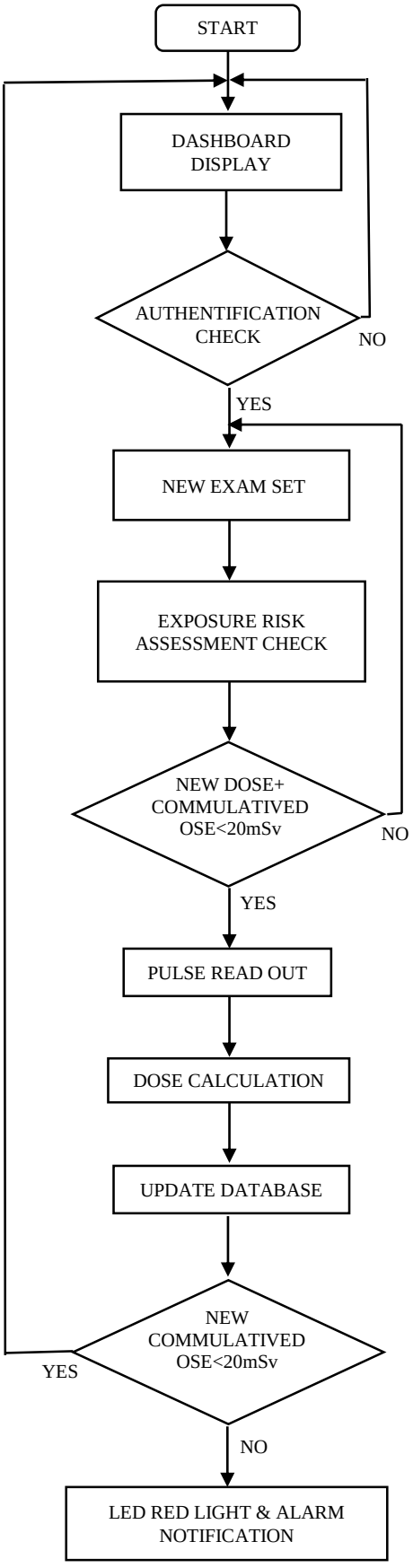


Fig. 5. Algorithm flowchart for radiation dose management system.

III. Results and Discussion

In diagnostic radiology, X-ray dosimetry is essential to optimize imaging procedures and minimize unnecessary radiation exposure to patients. This includes techniques such as computed tomography (CT), fluoroscopy, and conventional radiography. Concerning radiation in therapy, X-ray dosimetry plays a critical role in planning and delivering precise doses of radiation to cancerous tissues while sparing surrounding healthy tissues.

Exposure to prolonged or heightened levels of ionizing radiation elevates the risk of developing cancer, with the specific type and likelihood of cancer contingent upon the organs affected. Robust statistical evidence indicates an increased cancer incidence among populations exposed to substantial radiation doses, as exemplified by survivors of atomic bombings in Hiroshima and Nagasaki [37]. Additionally, elevated radiation doses have the potential to induce heritable genetic mutations. Acute exposure to high doses of radiation can precipitate Acute Radiation Syndrome, characterized by symptoms such as nausea, vomiting, diarrhea, and, in severe cases, damage to organs and tissues, with the severity directly proportional to the dosage.

Radiation protection agencies rely on sophisticated models to project the lifetime risk of cancer resulting from radiation exposure, accounting for variables such as age, gender, and radiation type.

An experiment utilizing the E7239GX Toshiba X-ray tube device was conducted to measure scattered radiation doses across different parameters, as outlined in Table III. This study was undertaken to evaluate the levels of scattered radiation generated under varied operational conditions, a critical factor in assessing radiation safety protocols and refining imaging techniques in medical environments. All measurements were recorded using our Smart X-ray Geiger Data Logger and cross-verified with the BR-6 Geiger counter device, as depicted in Fig. 6. This comprehensive approach ensures accurate data collection and validation, contributing to enhanced understanding and management of radiation exposure in clinical settings.

Fig. 6. Verification of scattered radiation measurements using smart X-ray Geiger Data Logger and BR-6 Geiger counter.

TABLE III. Parameters and Corresponding Scattered Radiation Doses Measured with E7239GX Toshiba X-ray Tube Device

High Voltage (kV)	Filament Current (mA)	Distance from the Source (m)	Dose for Scattered Radiation (uSv)
70	100	1	4.71
70	50	1	4.41
70	20	1	4.36
80	50	1	4.44
90	50	1	4.50
100	50	1	4.53
110	50	1	4.55
70	100	2	1.17
70	50	2	1.10
70	20	2	1.09
80	50	2	1.11
90	50	2	1.12
100	50	2	1.13
110	50	2	1.13

In summary, adjusting high voltage (kV) and filament current (mA) affects the quality and quantity of the X-rays produced, which in turn influence the dose for scattered radiation at different distances from the source. Lowering kV or mA can reduce radiation exposure but may affect image quality, while increasing distance from the source decreases radiation intensity but requires compensation with higher kV or mA settings to maintain diagnostic quality.

However, the amount of scattered radiation dose received can vary significantly depending on several key factors. These include the type of material through which the X-rays scatter, such as bone or soft tissue, and the angle at which scattering occurs. Distance from the radiation source plays a critical role as well, with intensity decreasing according to the inverse square law as distance increases. Effective shielding and room design are crucial in minimizing scattered radiation exposure, as they can absorb or redirect scattered X-rays away from personnel. Patient positioning also influences exposure levels, as different positions can direct scattered radiation towards different areas of the exam room. Additionally, factors related to X-ray equipment, such as tube voltage, current settings, and beam collimation, directly impact the amount of scattered radiation generated.

To manage the power supply, control the Geiger-Müller tube, and acquire measured dose values, we utilized the RadiationD V1.1 board. Fig. 7 illustrates the board's diagram, structured around several key components.

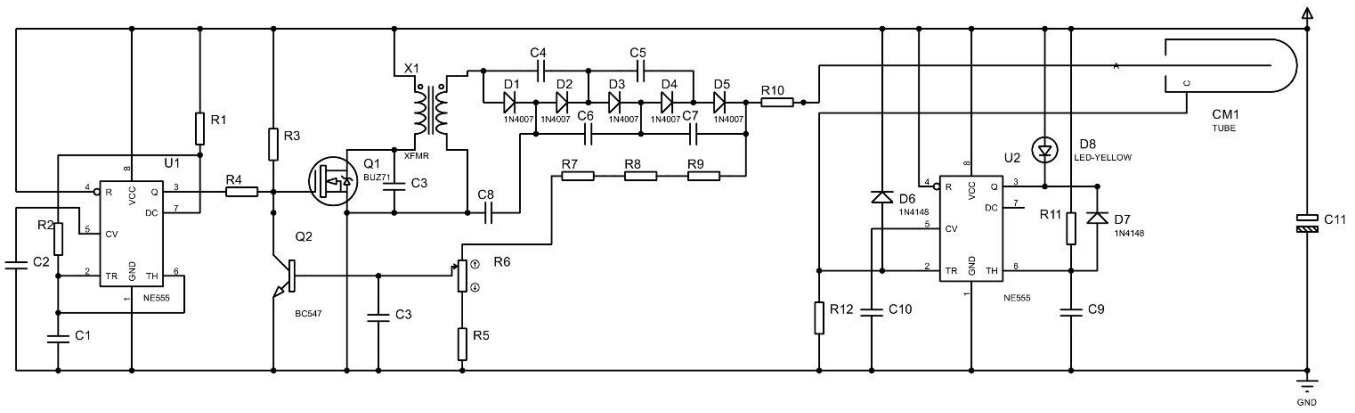


Fig. 7. RadiationD V1.1 board.

The first key component is the Square-Wave Generator Block. This block is tasked with generating a carrier signal that drives the system. The precise timing signals produced by this block ensure synchronization within the system, providing a stable foundation for the subsequent components to function correctly. Next, the High Voltage Power Supply plays a crucial role. This circuit converts the low voltage from the battery to the high voltage required to operate the Geiger-Müller tube, typically in the range of 350-500V. This component is essential for the stable operation of the Geiger-Müller tube, ensuring it receives the necessary high voltage to detect radiation accurately.

The Geiger-Müller Tube itself is a cylindrical tube filled with inert gas that becomes ionized when radiation passes through it. The ionization process generates electrical pulses, which are the primary signal used to detect radiation. This tube is the core of the radiation detection process, as it directly responds to ionizing particles. Capturing these pulses is the job of the Pulse Detection Circuit. This circuit typically consists of a resistor-capacitor (RC) configuration connected to the Geiger-Müller tube. It detects the electrical pulses generated by the tube when radiation is present. This detection circuit is crucial for converting the ionization events within the tube into readable electrical signals.

Finally, the Amplification and Timing Regulation component processes the detected pulses. This part of the board amplifies the pulses to ensure they are strong enough for accurate measurement. Additionally, it regulates the timing to ensure precise dose calculations. Accurate amplification and timing regulation are vital for providing reliable radiation dose values, ensuring that the detected signals are processed correctly.

Fig. 8 represents different amounts of X-ray radiation, depicted as a random signal. This figure illustrates the varying intensities and frequencies of X-ray radiation generated over a period of time, showcasing the fluctuating of radiation levels in the X-ray field.

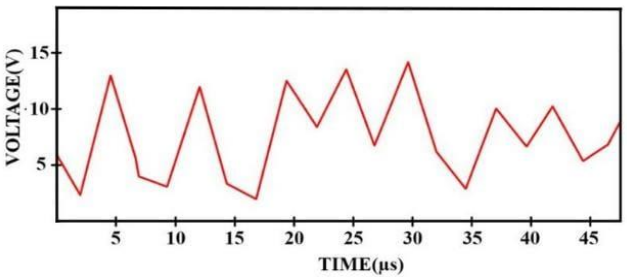


Fig. 8. Different amounts of x-ray radiation represented by a random signal.

Fig. 9 shows the high voltage generated for the Geiger-Müller tube's function. This figure highlights the conversion process from the low voltage power supply to the high voltage necessary for the operation of the Geiger-Müller tube. The stability and consistency of this high voltage are crucial for accurate radiation detection.

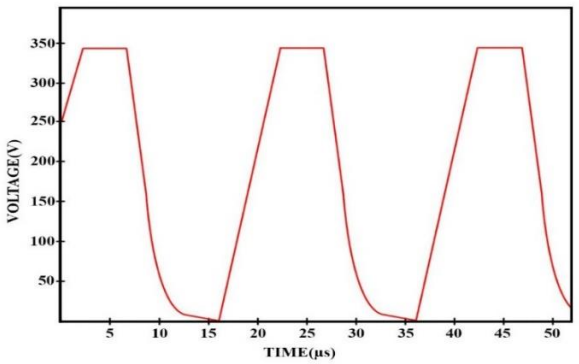


Fig. 9. High voltage generated for Geiger-Müller tube operation.

Fig. 10 represents the pulse detected by the Geiger-Müller tube, which will be transmitted for further processing. This figure demonstrates the output of the pulse detection circuit, showing the discrete electrical pulse generated in response to ionizing radiation passing through the Geiger-Müller tube. These pulses are then amplified and timed correctly for accurate radiation dose measurement and data analysis.

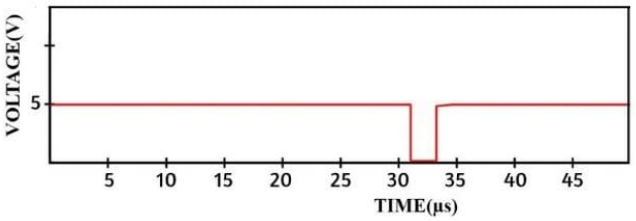


Fig. 10. Pulse detected by the Geiger-Müller tube for processing.

This study illustrates critical components and operational procedures essential for simulating X-ray doses. Among them, the RadiationD V1.1 board is prominently featured, serving as integral hardware for the detection system. These visual representations also depict different levels of X-ray radiation, demonstrating the system's sensitivity and response capabilities. Additionally, the figures demonstrate the generation of high voltage required for Geiger-Müller tube operation, underscoring its pivotal role in detecting radiation. Furthermore, they detail the pulses detected by the Geiger-Müller tube, providing insights into the electrical responses triggered by simulated radiation events. Together, this study offers a comprehensive overview of the experimental setup and operational mechanisms on simulating X-ray doses.

IV. Conclusion

In this study, we introduced the Smart X-ray Geiger Data Logger, a novel device designed for integrated detection, control, and dose evaluation of X-ray radiation. This device addresses the critical need for real-time monitoring and precise measurement of radiation doses in medical imaging and other applications. By leveraging the ionization effect within the Geiger-Müller tube and advanced algorithmic processing, the Smart X-ray Geiger Data Logger offers significant advancements in radiation safety management.

Through our experimental setup and data collection procedures, we demonstrated the device's capability to accurately measure and monitor X-ray doses, providing immediate feedback on both instant and cumulative radiation exposure. The integration of a robust algorithm for exposure risk assessment ensures proactive management of radiation safety, alerting users to potential risks and suggesting alternative procedures when safety limits are approached or exceeded. The Smart X-ray Geiger Data Logger stands poised to contribute significantly to the advancement of radiation safety practices, ensuring enhanced patient care and occupational safety across medical sectors.

Looking forward, the Smart X-ray Geiger Data Logger marks a significant advancement in radiation monitoring, providing healthcare professionals a vital tool for optimizing radiation exposure management.

References

[1] L.B. Youssef, A. Bybi, H. Drissi and E. A. Chater, "Enhancing Radiation Safety in Moroccan Healthcare: A Comprehensive Study on X-Ray Dose Monitoring, Control, and Dosimeter Integration," 2024 4th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), FEZ, Morocco, 2024, pp. 1-6, doi: 10.1109/IRASET60544.2024.10548161.

[2] Eaton, D. J., Gonzalez, R., Duck, S., & Keshtgar, M. (2011). Radiation protection for an intra-operative X-ray device. The British journal of radiology, 84(1007), 1034-1039.

[3] Acerbi, F., Altamura, A. R., Di Ruzza, B., Merzi, S., Spinnato, P., & Gola, A. (2023). Characterization of radiation damages on Silicon photomultipliers by X-rays up to 100 kGy. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 1045, 167502.

[4] L.B. Youssef, A. Bybi, H. Drissi, and E.A. Chater, "Dose Archiving and Communication System in Moroccan Healthcare: A Unified Approach to X-Ray Dose Management and Analysis," International Journal of Advanced Computer Science and Applications, vol. 15, no. 8, pp. 590-601, 2024.

[5] Fum, W. K. S., Wong, J. H. D., & Tan, L. K. (2021). Monte Carlo-based patient internal dosimetry in fluoroscopy-guided interventional procedures: A review. Physica Medica, 84, 228–240. https://doi.org/10.1016/j.ejmp.2021.03.004.

[6] Yang, S., Zhang, X., Deng, C., Zhang, Y., Xu, F., Guan, L., & Duan, S. (2022). Design of portable multi-function radiation detection system. He Jishu/Nuclear Techniques, 45(11). https://doi.org/10.11889/J.0253-3219.2022.HJS.45.110403.

[7] Johnson, R., & Patel, K. (2022). Novel techniques in X-ray dose detection: A comprehensive review. International Journal of Medical Physics, 38(2), 157-172.

[8] Garcia, M., & Kim, H. (2021). Impact of automated X-ray dose tracking systems on patient safety. Radiation Protection Dosimetry, 50(4), 303-317.

[9] Brown, L**.,** & Nguyen, T. (2020). Enhancing X-ray imaging safety with advanced dosimetry methods. Medical Imaging Journal, 32(1), 45-59.

[10] Thomas, J., & Zhao, L. (2022). Reducing radiation exposure: Innovations in X-ray dose monitoring technologies. Journal of Medical Imaging and Radiation Sciences, *44*(2), 89-103.

[11] Wilson, P., & Martinez, D. (2019). Development of portable X-ray dose detectors for clinical use. *Clinical Radiology Advances*, *29*(5), 401-415.

[12] Smith, J., & Lee, A. (2023). Advancements in real-time X-ray dose monitoring for radiological procedures. Journal of Radiological Science and Technology, *45*(3), 211-225.

[13] Robinson, K., & Wang, X. (2021). Real-time X-ray dose management in diagnostic imaging: Challenges and solutions. Journal of Diagnostic Radiology, 39(4), 221-235.

[14] Martinez, R., & Lee, S. (2019). Application of AI in X-ray dose detection and monitoring. *Artificial Intelligence in Medicine*, *47*(1), 65-79.

[15] Silva, W. R. F., & Fonseca, J. M. (2023). Development of an X-ray machine, a Geiger detector and a goniometer of low-cost to study X-rays. Revista Brasileira de Ensino de Fisica, 45. https://doi.org/10.1590/1806-9126-RBEF-2023-0073.

[16] International Commission on Radiological Protection (ICRP). (2007). The 2007 Recommendations of the International Commission on Radiological Protection. Annals of the ICRP, 37(2-4), 1-332.

[17] Grasso, S., Varallo, A., Ricciardi, R., Italiano, M. E., Oliviero, C., D'Avino, V., Feoli, C., Ambrosino, F., Pugliese, M., & Clemente, S. (2023). Absorbed dose evaluation of a blood irradiator with alanine, TLD-100 and ionization chamber. Applied Radiation and Isotopes, 200. https://doi.org/10.1016/j.apradiso.2023.110981.

[18] Waqar, M., Ul-Haq, A., Bilal, S., & Masood, M. (2017). Comparison of dosimeter response of TLD-100 and ionization chamber for high energy photon beams at KIRAN Karachi in Pakistan. Egyptian Journal of Radiology and Nuclear Medicine, 48(2), 479–483. https://doi.org/10.1016/j.ejrnm.2017.01.012.

[19] Neves, L. P., Perini, A. P., Fernández-Varea, J. M., Cassola, V. F., Kramer, R., Khoury, H. J., & Caldas, L. V. E. (2014). Dosimetric application of a special pencil ionization chamber in radiotherapy X-ray beams. Radiation Physics and Chemistry, 95, 98–100. https://doi.org/10.1016/j.radphyschem.2012.12.042.

[20] Zhou, M., Hu, L., Huang, L., Zhong, G., Li, K., Hong, B., Xiao, M., & Zhang, R. (2020). Measurement of the radiation dose distribution in EAST hall based on thermoluminescence dosimeter. Fusion Engineering and Design, 160, 111977. https://doi.org/10.1016/J.FUSENGDES.2020.111977.

[21] Del Sol Fernández, S., García-Salcedo, R., Sánchez-Guzmán, D., Ramírez-Rodríguez, G., Gaona, E., de León-Alfaro, M. A., & Rivera-Montalvo, T. (2016). Thermoluminescent dosimeters for low dose X-ray measurements. Applied Radiation and Isotopes, 107, 340–345. https://doi.org/10.1016/J.APRADISO.2015.11.021.

[22] Ghoneam, S. M., Mahmoud, K. R., Diab, H. M., & El-Sersy, A. (2022). Studying the dose level for different X-ray energy conventional radiography by TLD-100. Applied Radiation and Isotopes, 181, 110066. https://doi.org/10.1016/J.APRADISO.2021.110066.

[23] Moradi, F., Mahdiraji, G. A., Rezaee Ebrahim Saraee, K., Khandaker, M. U., Adikan, F. R. M., & Bradley, D. A. (2022). Impact of dosimeter size on energy dependence: An experimental study on glass TLDs. Radiation Physics and Chemistry, 200, 110176. https://doi.org/10.1016/J.RADPHYSCHEM.2022.110176.

[24] Bakkari, M., & Soliman, K. (2018). [P004] Measurment of entrance surface dose during chest X-ray examinations in neonatal intensive care unit using OSL and TLD dosimeters. Physica Medica, 52, 100. https://doi.org/10.1016/J.EJMP.2018.06.337.

[25] Thabit, H. A., Ismail, A. K., Kabir, N. A., al Mutairi, A. M., Bafaqeer, A., Alraddadi, S., Jaji, N. D., Sayyed, M. I., & Al-Ameri, S. M. (2023). Investigation of the thermoluminescence dosimeter characteristics of multilayer ZnO(300 nm)/Ag(50 nm)/ZnO(x) thin films for photonic dosimetry applications. Optical Materials, 137, 113548. https://doi.org/10.1016/J.OPTMAT.2023.113548.

[26] Tang, K., & Zhang, S. (2021). Real-time dosimeter based on LiF:Mg,Cu,P and SiPM. Radiation Measurements, 145, 106607. https://doi.org/10.1016/J.RADMEAS.2021.106607.

[27] Posar, J. A., Davis, J., Brace, O., Sellin, P., Griffith, M. J., Dhez, O., Wilkinson, D., Lerch, M. L. F., Rosenfeld, A., & Petasecca, M. (2020). Characterization of a plastic dosimeter based on organic semiconductor photodiodes and scintillator. Physics and Imaging in Radiation Oncology, 14, 48–52. https://doi.org/10.1016/J.PHRO.2020.05.007.

[28] Pejović, M. M., & Pejović, S. M. (2023). P-channel MOSFET as ionizing radiation detector. Applied Radiation and Isotopes, 196. https://doi.org/10.1016/j.apradiso.2023.110730.

[29] Matsumoto, T., Yamaguchi, K., Yanagisawa, R., Kubodera, K., Arai, Y., Makino, T., Ohshima, T., Sakai, M., Matsumura, A., & Kada, W. (2023). Development of a SiC semiconductor-based dosimeter for evaluating clinical dose distribution in carbon ion cancer therapy fields. Nuclear Instruments and Methods in Physics Research, Section B: Beam Interactions with Materials and Atoms, 542, 151–157. https://doi.org/10.1016/j.nimb.2023.06.013.

[30] Jung, H., Lee, K. J., Kim, J. L., & Lee, S. Y. (2004). Development of a personal dosimeter badge system using sintered LiF:Mg,Cu,Na,Si TL detectors for photon fields. Radiation Measurements, 38(1), 71–80. https://doi.org/10.1016/j.radmeas.2003.08.006.

[31] Yamadera, A., Kim, E., Miyata, T., & Nakamura, T. (1995). Development of high sensitivity X-and λ-ray personal dosimeter using photostimulated luminescent detector. Applied Radiation and Isotopes, 46(6–7), 467–468. https://doi.org/10.1016/0969-8043(95)00053-4.

[32] Alnawaf, H., Butson, M. J., Yu, P. K. N., & Cheung, T. (2011). SIRAD – Personal radiation detectors. Radiation Measurements, 46(12), 1826–1828. https://doi.org/10.1016/J.RADMEAS.2011.07.027.

[33] Guo, C. Y., Lin, T. L., & Hsieh, T. L. (2022). A Solar-Rechargeable Radiation Dosimeter Design for Radiation Hazard Zone Located with LoRa Network. Quantum Beam Sci. 2022, 6, 27.

[34] Babiuch, M., Foltýnek, P., & Smutný, P. (2019, May). Using the ESP32 microcontroller for data processing. In 2019 20th International Carpathian Control Conference (ICCC) (pp. 1-6). IEEE.

[35] Osman, H., Ahmed, A. M., Musa, A., Medani, A., Abouraida, R. A., Alelyani, M., ... & Awadallah, B. A. (2024). Radiation dose assessment: Establishment of local diagnostic reference levels for selected radiography examinations across three prominent hospitals in Sudan. Radiation Physics and Chemistry, 217, 111482.

[36] M. Alkhorayef, A. Sulieman, Khalid Alzahrani, Mohamed Abuzaid, Othman I. Alomair, M. Almuwannis, Salem Alghamdi, Nissren Tamam, David A. Bradley, Radiation risk for patients undergoing cardiac computed tomography examinations, Applied Radiation and Isotopes, Volume 168, 2021, 109520, ISSN 0969-8043, https://doi.org/10.1016/j.apradiso.2020.109520.

[37] Chun, C. K. S. (2008). *Japan 1945:* From Operation Downfall to Hiroshima and Nagasaki. Osprey Publishing.

A Feature Map Adversarial Attack Against Vision Transformers

Majed Altoub[1], Rashid Mehmood[2], Fahad AlQurashi[3], Saad Alqahtany[4], Bassma Alsulami[5]
Department of Computer Science, Faculty of Computing and Information Technology
King Abdulaziz University, Jeddah 21589, Saudi Arabia[1,3,5]
Department of Computer Science, Faculty of Computer and Information Systems
Islamic University of Madinah, Madinah 42351, Saudi Arabia[2,4]

***Abstract*—Image classification is a domain where Deep Neural Networks (DNNs) have demonstrated remarkable achievements. Recently, Vision Transformers (ViTs) have shown potential in handling large-scale image classification challenges by efficiently scaling to higher resolutions and accommodating larger input sizes compared to traditional Convolutional Neural Networks (CNNs). However, in the context of adversarial attacks, ViTs are still considered vulnerable. Feature maps serve as the foundation for representing and extracting meaningful information from images. While CNNs excel at capturing local features and spatial relationships, ViTs are better at understanding global context and long-range dependencies. This paper proposes a feature map ViT-specific adversarial example attack called Feature Map ViT-specific Attack (FMViTA). The objective of the investigation is to generate adversarial perturbations in the spatial and frequency domains of the image representation that allow deeper distance measurement between perturbed and targeted images. The experiments focus on a ViT pre-trained model that is fine-tuned on the ImageNet dataset. The proposed attack demonstrates the vulnerability of ViTs to adversarial examples by showing that even allowing only 0.02 maximum perturbation magnitude to be added to the input samples gives 100% attack success rate.**

Keywords*—Vision transformers; adversarial attacks; DNNs; vulnerabilities; feature maps; perturbations; spatial domains; frequency domains

I. Introduction

Deep Neural Networks (DNNs) have emerged as highly effective image classification tools. Convolutional Neural Networks (CNNs) recognize the input image data by the convolutional layers to identify and capture local patterns and spatial hierarchies of features [1]. On the other hand, Vision Transformers (ViTs) represent the input image data as sequences of patches and leverage the self-attention mechanisms to capture long-range dependencies and global context in images [2]. Therefore, ViTs can be more effective than CNNs in long-range interactions and achieve state-of-the-art performance on various image classification benchmarks.

Feature maps are fundamental for representing and extracting valuable information from images in DNN classification models. ViTs are better at understanding global context and long-range dependencies than CNNs [3]. ViTs divide the input image into patches and then flatten and embed them into tensors. These tensors are then fed into a transformer encoder, which learns to attend to different patches and their relationships.

From the security perspective, the adversarial robustness of the ViT models is a significant challenge, specifically in a domain such as image classification. Adversarial attacks intentionally manipulate the models by inserting small perturbations into the input image data to deceive them, resulting in incorrect predictions of the input image while appearing normal to humans, called adversarial example attacks. Adversarial perturbations are critical in understanding the vulnerabilities of the ViT models to adversarial examples.

The existent studies of adversarial perturbations have focused on two main areas: the spatial domain and the frequency domain [4]. In the spatial domain, the representation of the input images is extracted directly from the pixel values which is typically used in attacks such as gradient-based adversarial attacks [5]–[10]. The frequency domain including low-frequency and high-frequency is another perspective that can be used to generate perturbations in selected frequency regions [4]. The representation of the input images in the frequency domain is transformed through methods like the Discrete Cosine Transform (DCT) [11]. Prior research has indicated that CNNs are vulnerable to high-frequency noise and ViTs are vulnerable to low-frequency noise [12]. These findings have contributed to the emergence of frequency-based adversarial attacks, including [11], [13]–[17].

In this research paper, our objective is to investigate the potential impact of introducing an adversarial example attack by generating the adversarial perturbations in the spatial domain but with some potential influence from the frequency domain. We introduce the FMViTA attack that uses the ViT feature maps of the perturbed and the target images to optimize the adversarial perturbations. For deeper comparison, the cosine similarity loss function is used to measure the similarity between feature maps. The ViT model itself can learn to capture frequency-related information through its self-attention mechanism. This means that the feature maps used for cosine similarity might contain some frequency-based representations, especially since we take the mean feature maps from the intermediate blocks. While the ViT feature maps are not considered directly spatial domains, they are still influenced by the spatial arrangement of the image patches. However, adding adversarial perturbations to the input image is a direct pixel-level operation happening in the spatial domain. Furthermore, for imperceptibly added, we clamp adversarial perturbations.

We, however, do not intend to make a strength transferable adversarial example attack; we pose a direction of generating perturbation by using more deeper comparison methods to optimize the perturbations generated. Our results demonstrate

that using FMViTA method can indeed manipulate a ViT model to classify a set of five input images as other target images with 100% attack success rate, while we only allowing 0.02 maximum perturbation magnitude to be changed. The contributions of this paper are summarized as follows:

- We employ a cosine similarity loss function for deeply measuring the similarity between the perturbated and the target feature maps of the images to optimize the perturbation patterns.
- We also propose a novel Feature Map ViT-specific Attack method named FMViTA that takes one input image and runs it against a target image to generate a perturbated image that looks like the input one but classifies as the target image.
- We demonstrate the results of using FMViTA against a ViT for five input and target images.

The rest of the paper is organized as follows: Section II briefly gives a background and reviews the related works. In Section III, we provide our methodology, including the threat model. Experiments setup and results are presented in Section IV. Finally, the conclusion is made in Section V.

II. Background and Related Work

A. Adversarial Attacks in DNNs

The adversarial attacks can be broadly classified by the attack goals to evasion, poisoning, backdoor, and privacy attacks. In evasion attacks, the goal is to cause misclassification at test time, such as in the adversarial example attacks [18]. In the poisoning attacks, the goal is to inject poisonous data that can manipulate the DNN model's training. The goal of the privacy attacks is to steal information from the DNN models. Backdoor attacks can be considered both evasion and poisoning attacks because adversaries inject poisonous data called triggers into legitimate DNN models at training time and misclassify the model once the trigger is activated at test time [19].

From another point of view, adversarial attacks can be classified based on the accessibility of the adversaries. In this context, adversarial attacks can be either white-box or black-box attacks. White-box attacks are powerful because adversaries can access the architecture and parameters of a DNN model. On the other hand, black-box attacks are more challenging because the adversaries can only access the input and output data from a DNN model. Gray-box attacks are situated between white-box and black-box attacks. In these attacks, the adversaries possess limited knowledge of the model's architecture and parameters.

In addition, depending on the goals of the attacks, adversarial attacks can be classified as targeted and untargeted. In untargeted attacks, the goal is to misclassify the DNN classification model to output any other class. In the targeted attacks, the goal is to misclassify the DNN classification model to classify the input as a particular class. In this paper, our attack method is considered a targeted gray-box adversarial example attack. However, the method could be developed to generate other types of attacks.

B. Adversarial Example Attack Methods

The earlier adversarial example attacking phenomena were proposed by Szegedy et al. [18]. Since then, various methods for creating highly powerful perturbations have been put forward in academic literature. Goodfellow et al. [5] first introduced the concept of adversarial examples using a gradient-based method called the Fast Gradient Sign Method (FGSM). In this method, the gradient of the loss function is used with respect to the input data to generate the small perturbation that maximizes the loss function, which can be conceded as a white-box attack. On the other hand, Kurakin et al. [6] improve the FGSM by running multiple iterations of the attack. However, the method showed that there would still be a gap in accuracy between the perturbed results and the being ones. Later researchers use momentum [7] approaches to further enhance the adversarial example attacks. Dong et al. [8] proposed the Momentum Iterative Fast Gradient Sign Method (MI-FGSM) to increase the success rate of the black-box attacks, which improves the transferability of the adversarial example attacks. Other enhancements were proposed as well, such as the Ensemble Momentum Iterative FGSM (EMI-FGSM) [9] and VMI-FGSM and VNI-FGSM [10].

Frequency-based adversarial example attacks have been developed in the literature to generate adversarial perturbations. Recently, Duan et al. [11] proposed another way of crafting adversarial examples named AdvDrop. Instead of adding perturbations, they drop existing details from clean images from frequency components. The results demonstrate that AdvDrop can achieve high attack success rates on ImageNet dataset. Guo et al. [15] focus on the low frequency adversarial perturbation for black-box attacks. The authors demonstrate the effectiveness of this technique by successfully fooling the Google Cloud Vision platform with an unprecedented low number of only 1000 model queries . Jin et al. [17] introduced a novel white-box attack known as the Frequency and Spatial Consistency Based Adversarial Attack (FSA). The results show that the FSA method can enhance the success attack rates, with a maximum improvement of 28.98% observed across various attack methods and models.

C. Adversarial Robustness of Vision Transformer

ViT is state-of-the-art for image classification that uses a transformer architecture to process image patches and generate predictions [20]. As ViTs continue to gain popularity and are being deployed in various domains, the robustness of ViTs to adversarial attacks becomes essential.

According to Aldahdooh et al. [21] Vanilla ViTs or hybrid ViTs are more robust to adversarial attacks than CNNs, particularly under Lp-norm or adaptive attacks. The findings demonstrate that increasing the number of attention blocks may increase the robustness to transfer attacks but not white-box attacks. Shao et al. [22] claim that ViTs are less sensitive to high-frequency perturbations than CNNs and MLP-Mixers. That is because ViTs contain less high-frequency features, which have a high correlation with their robustness against different frequency-based perturbations. In addition, in ViTs, introducing convolutional or token-to-token blocks for learning high-frequency features can improve classification accuracy, but at the cost of adversarial robustness. Furthermore, modern

CNNs may borrow some of ViTs techniques to bridge the performance gap as well as the adversarial robustness gap.

Paul and Chen [23] have compared the robustness of ViTs and CNNs models against frequent corruptions and perturbations using six different ImageNet datasets. Their findings suggest that ViTs have better robustness performance than CNNs on some of the datasets, and that the robustness of both models can be improved by fine-tuning on corrupted data. Overall, the study highlights the potential of ViTs as a promising alternative to CNNs in computer vision tasks, especially when it comes to robustness. Joshi et al. [24] analyzed the underlying distinctions between the ViT and CNN models. Unlike convolutional networks, vision transformers use a patch token-based self-attention mechanism. By creating a block sparsity based adversarial token attack, their study found that ViT models are more sensitive to token attacks than CNN models. This highlights the need for robustness evaluation of transformer-based models against adversarial attacks. It also suggests that further research is required to improve the security of vision transformers.

III. Methodology

A. Threat Model

We use the threat model to define our FMViTA attack scenario. We focus on image classification; nevertheless, the technique can be easily extended to other domains. In this threat model, we assume that the adversary has limited knowledge of the model's architecture and parameters. In addition, the adversary does not have access to the training data. The attack is considered a gray-box adversarial attack. The adversary's objective in this threat model is to embed perturbations into the input images to make them perturbed in order to fool the ViT classification model to classify them as target images' classes.

B. Attack Method

We formulate FMViTA as an optimization problem to generate perturbed images. The optimization is based on minimizing the distance between feature representations of the perturbed and target images. To do so, we take the mean of the intermediate features for both input and target images. Then we calculate the cosine similarity between the mean features of the perturbed and target images to measure the distance between them, which is the loss function to be used in the gradient iterations.

The loss function encourages the features of the perturbed image to be similar to those of the target image. Our attack utilizes the intermediate feature maps instead of focusing on the model's output probabilities. Thus, we chose the cosine similarity as a function due to its suitability for comparing high-dimensional feature vectors. Furthermore, the cosine similarity concentrates on the angular relationship between the two feature maps. Compared to other similarity calculations, such as Euclidean and Manhattan distances, cosine similarity is less sensitive to magnitude differences because it focuses on the angle between vectors. This property allows for a more in-depth comparison between the two feature maps. Since the cosine similarity is used to measure this similarity, a higher similarity results in a lower loss as shown in Eq. (1).

$$L(F_t, F_p) = 1 - \frac{F_t \cdot F_p}{\|F_t\|\|F_p\|} \tag{1}$$

$L(F_t, F_p)$: The loss function measuring the difference between the feature representations of the target image F_t and the perturbed image F_p.

$F_t \cdot F_p$: The dot product of the feature vectors, representing their similarity.

$\|F_t\|$and$\|F_p\|$: The magnitudes (norms) of the feature vectors, normalizing the cosine similarity.

Let M be the target ViT model and I_t and I_s be the input and target images. The optimizing perturbed patterns P is added to the input image as to generate perturbed images I_p as shown in Eq. (2). In addition, we clamp the perturbed pattern values to prevent them from becoming too large as shown in Eq. (3).

The gradient optimization can be mathematically described as:

$$\nabla_{t_i} L_i = \nabla_{t_i} \left(1 - L\left(mean\left(F_p\left(I_p\right)\right), mean(F_t(I_t))\right)\right) \tag{2}$$

$$I_p = \text{clamp}(I_s + P, -0.02, 0.02) \tag{3}$$

Algorithm 1: Feature Map ViT-specific Attack (FMViTA)

Input: Target model M, target image I_t, input images I_s, perturbed pattern P. **Output:** Perturbed images I_p.

Initialize P zeros.

for *each epoch* e **do**

  for *each input image* I_s *in* I_s **do**

    $I_p \leftarrow \text{Clamp}(I_s + P)$ Generate perturbed image

    $F_t \leftarrow M(I_t)$ Get features of target image

    $F_p \leftarrow M(I_p)$ Get features of perturbed image

    $L \leftarrow \text{Loss}(F_t, F_p)$ Calculate loss

    Update P using gradient descent: $P \leftarrow P - \alpha \nabla L$

  Return: Perturbed images I_p

IV. Experiments and Results

In this section, we first explain our experimental setup and then show the experimental results.

A. Experimental Setup

Our experiments mainly focus on pre-trained ViT (google/vit-base-patch16-224) [25], [26] that was pre-trained on ImageNet-21k [27] and fine-tuned on ImageNet [26] with a resolution of 224x224 and 1000 classes. Instead of using Adam, we use AdamW optimizer with a learning rate of 0.06 and weight decay of 0.002. The experiments were conducted on a dataset consisting of various AI-generated images for the input and target images that are classified as the ImageNet labels. We use AdamW optimizer to add the regularization

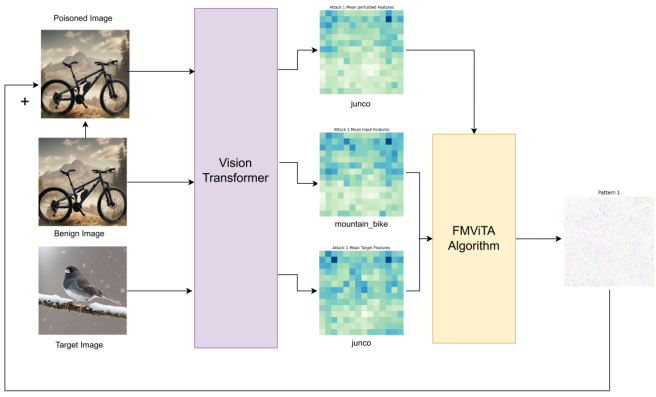


Fig. 1. The overall view of FMViTA attack ruining for Attack 1.

benefits of L2 regularization. We use AI-generated images as a dataset samples to be the target, and for the input images, we use five classes from the ImageNet dataset, which are {junco, mountain bike, terrapin, soccer ball, and goldfinch}.

Adam is an algorithm for first-order gradient-based optimization that computes adaptive learning rates for different parameters by maintaining estimates of the first and second moments of the gradients [28]. Adam combines the advantages of AdaGrad and RMSProp, with key differences in how it updates parameters and handles bias correction. However, L2 regularization and weight decay are equivalent for standard stochastic gradient descent (SGD) [29], but not for adaptive gradient algorithms like Adam. AdamW [30] modifies the popular Adam optimizer to improve its generalization performance, allowing it to compete with SGD with momentum on image classification tasks. To take advantage of AdamW we use it for the gradient-based optimization of FMViTA attack.

B. Experimental Results

We run the experiments to test the effectiveness of our attack algorithm using the same adjusted hyperparameters, such as the learning rate of 0.06, weight decay of 0.002, and 700 epochs for all five attacks. These hyperparameters were determined through an iterative process of empirical evaluation. The choice of a 0.06 learning rate suggests a balance between speed and stability for all five attacks to be run as a one-time attack; however, each individual attack can be run using a different learning rate based on the images being used. The weight decay parameter incorporated in the AdamW optimizer offers L2 regularization to address overfitting, consequently improving the generalization of the generated adversarial examples. It was set at 0.002 to provide the best balance between preventing overfitting and maintaining sufficient subtle adversarial perturbations. We decided to use 700 epochs after monitoring the convergence behaviors of the attack's loss functions.

In order to assess the effectiveness of the proposed FMViTA attack, we conducted an evaluation of the attack's capability to manipulate the source of five different input images and five corresponding target images. This paired selection allowed a more robust evaluation, and to determine how well the attack generalizes to unseen data. In Attack 1, we use a mountain bike as an input image, then after extracting the feature maps from the 12 blocks of the ViT model, we take the mean of all the maps to be in one tensor, and we do the same for the target image, which is a junco. Then we run the attack algorithm using cosine similarity as the loss function between the mean feature map of the perturbed image (the input image + perturbed pattern) and the mean feature map of the target image. We optimize the perturbed pattern using the AdamW optimizer to add the regularization benefits of L2. The goal is to make the mean feature map of the perturbed image as close as the mean feature map of the target image. The overall illustration of Attack 1 can be seen in the Fig. 1.

In Attack 2, we do it the opposite way, using a "junco" as an input image and a "mountain bike" as the target image. Then, we do the same as in Attack 1. In Attack 3 the input image was "soccer ball" and the target image was a "terrapin". In Attacks 4 and 5 "goldfinch" and "terrapin" were the input images, and "soccer ball" and "goldfinch" were the target images. All attacks' results are shown in Fig. 3.

To evaluate the attack performance, we use the attack success rate (ASR) to measure the success of the FMViTA attack. The ASR refers to the percentage of FMViTA that effectively induces a misclassification in the ViT model, and it can be represented as shown in Eq. (4).

$$ASR_{\text{FMViTA}} = \frac{N_{\text{mis}}}{N_{\text{t}}} * 100 \tag{4}$$

where: N_{mis} is the number of perturbed images from FMViTA that are misclassified by the ViT model. N_{t} is the total number of experiments.

In order to enhance the existing evaluation method, we have fixed the parameters for a group of five input images against the other five target images. Our results show that the attack achieved a success rate of 100%. However, in terms of the invisibility of the noise in perturbed images, the attack result can be better when using only one input image against one target image and adjusting parameters based on that.

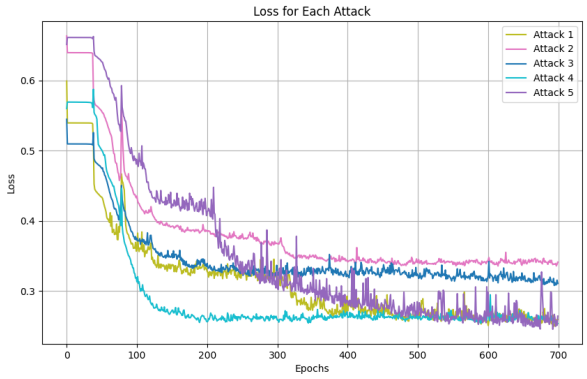


Fig. 2. The loss function results of the 5 attacks with 700 iterations.

That can be seen in Fig. 2 were losses behaved differently in each single attack iteration. Also, we found that extending the epochs did not improve the accuracy of in terms of the

invisibility of the noise in perturbed images, as if we changed the other parameters. The FMViTA attack tool will be available online with the flexibility to change the attack parameters.

Fig. 3. The results of FMViTA for a group of 5 input images against the other 5 target images.

Adversarial patterns often used in adversarial attacks, are specific inputs designed to elicit certain responses from a model. These patterns can expose the internal workings of neural networks and provide insights into their behavior under various conditions. Our patterns are typically generated through the FMViTA method to manipulate the input images intentionally.That can be seen in Fig. 4.

FMViTA method relies on feature maps, but the process of extracting features is computationally intensive, especially for long epochs and iterations in such an attack. Our optimization strategies are crucial for making the attack feasible within a reasonable timeframe. In FMViTA we extract the mean feature maps of the input and target images only once. However, it is necessary to extract the mean feature map of the perturbed images in each iteration. Fig. 5 shows the differences in mean feature maps of the input, target, and perturbed images.

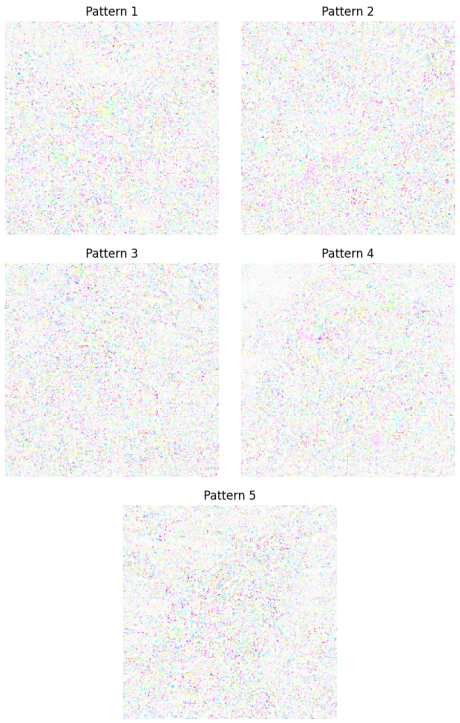


Fig. 4. The attack patterns for Attacks 1-5.

The vulnerability of ViTs to FMViTA stems from the exploitation of the intermediate feature representations within the DNN. The attack's efficacy derives from targeted manipulation of the intermediate features of the input image, leveraging gradient-based optimization, to generate perturbations that maximize the similarity between these intermediate features and the corresponding intermediate features of the target image. Thus, the success of FMViTA highlights the susceptibility of ViTs to gradient-based attacks targeting intermediate representations. Therefore, robust architectures and defenses against such sophisticated manipulation techniques are needed.

Mitigating the effectiveness of this attack approach suggests focusing on feature-space defenses. That involves analyzing the intermediate feature representations within ViTs to detect or mitigate the adversarial perturbations. One possible strategy is to apply denoising techniques to the intermediate feature maps. Recently, Yang et al. [31] have proposed the Denoising Vision Transformers (DVT) to suppress the grid-like artifacts observed in the feature maps by separating the clean features from those contaminated and then train a lightweight transformer block to predict clean features from raw ViT outputs. Another approach to mitigate such an attack is modifying the attention mechanism itself to enhance the robustness of attention weights to adversarial influence. That involves incorporating regularization terms during training or employing attention-aware denoising.

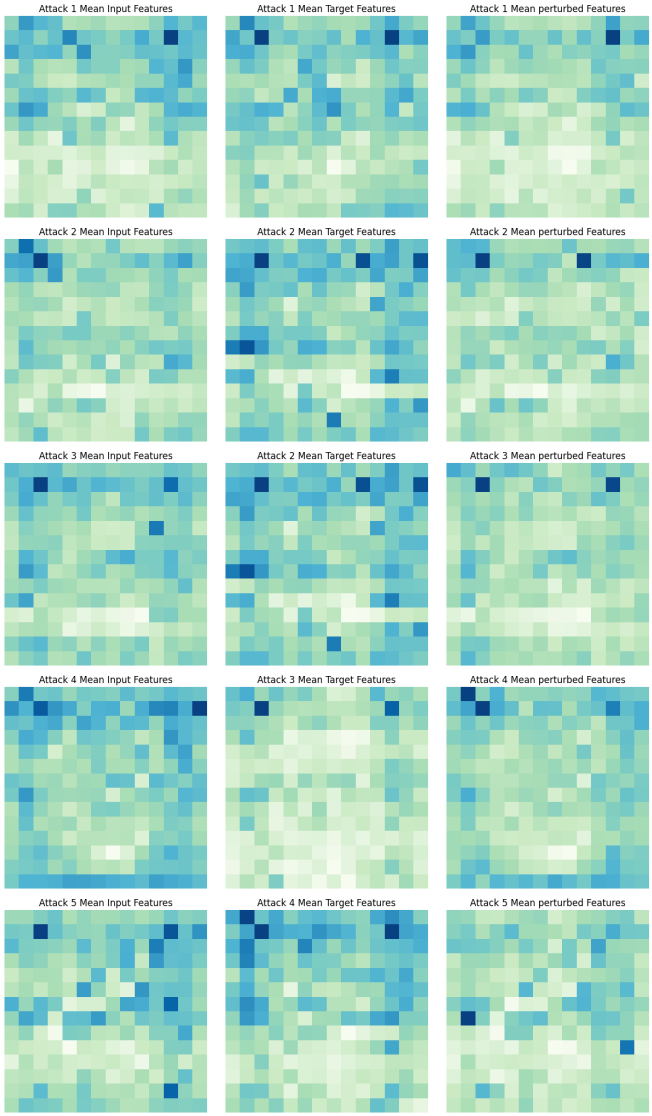


Fig. 5. The mean feature maps of the input, target, and perturbed images for Attacks 1-5.

V. Conclusion

The study found that vision transformers are vulnerable to FMViTA, which can significantly degrade their output accuracy. Therefore, developing more robust ViTs is crucial for ensuring the accuracy and reliability of ViT models. This highlights the importance of developing robust defense mechanisms to mitigate the impact of such attacks on vision transformers. However, investigating ViTs' robustness by generating adversarial attacks is a key to crafting robust defense mechanisms.

For future work, we will develop this direction of generating adversarial perturbations to be more robust to defense mechanisms. Furthermore, we will continue improving the FMViTA's performance in comparison with other state-of-the-art attacks. In addition, transferable improvements in the attack to be run in other ViT models are essential. Further, we can extend the FMViTA method to create other types of adversarial attacks, such as backdoor and poisoning attacks.

Acknowledgment

This article is derived from a research grant funded by the Research, Development, and Innovation Authority (RDIA), Kingdom of Saudi Arabia, with grant number 12615-iu-2023-IU-R-2-1-EI-.

References

[1] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," *arXiv e-prints*, p. arXiv:1511.08458, Nov. 2015.

[2] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *arXiv e-prints*, p. arXiv:2010.11929, Oct. 2020.

[3] J. Maurício, I. Domingues, and J. Bernardino, "Comparing vision transformers and convolutional neural networks for image classification: A literature review," *Applied Sciences*, vol. 13, no. 9, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/9/5521

[4] G. Kim, J. Kim, and J. Lee, "Exploring adversarial robustness of vision transformers in the spectral perspective," in *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Los Alamitos, CA, USA: IEEE Computer Society, jan 2024, pp. 3964–3973. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/WACV57701.2024.00393

[5] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and Harnessing Adversarial Examples," *arXiv e-prints*, p. arXiv:1412.6572, Dec. 2014.

[6] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial Machine Learning at Scale," *arXiv e-prints*, p. arXiv:1611.01236, Nov. 2016.

[7] B. Polyak, "Some methods of speeding up the convergence of iteration methods," *USSR Computational Mathematics and Mathematical Physics*, vol. 4, no. 5, pp. 1–17, 1964. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0041555364901375

[8] Y. Dong, F. Liao, T. Pang, H. Su, J. Zhu, X. Hu, and J. Li, "Boosting Adversarial Attacks with Momentum," *arXiv e-prints*, p. arXiv:1710.06081, Oct. 2017.

[9] X. Wang, J. Lin, H. Hu, J. Wang, and K. He, "Boosting adversarial transferability through enhanced momentum," in *British Machine Vision Conference*, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:232290454

[10] X. Wang and K. He, "Enhancing the transferability of adversarial attacks through variance tuning," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 1924–1933.

[11] R. Duan, Y. Chen, D. Niu, Y. Yang, A. K. Qin, and Y. He, "Advdrop: Adversarial attack to dnns by dropping information," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 7486–7495.

[12] N. Park and S. Kim, "How Do Vision Transformers Work?" *arXiv e-prints*, p. arXiv:2202.06709, Feb. 2022.

[13] H. Wang, X. Wu, Z. Huang, and E. P. Xing, "High-frequency component helps explain the generalization of convolutional neural networks," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8681–8691.

[14] D. Yin, R. G. Lopes, J. Shlens, E. D. Cubuk, and J. Gilmer, *A fourier perspective on model robustness in computer vision*. Red Hook, NY, USA: Curran Associates Inc., 2019.

[15] C. Guo, J. S. Frank, and K. Q. Weinberger, "Low Frequency Adversarial Perturbation," *arXiv e-prints*, p. arXiv:1809.08758, Sep. 2018.

[16] Y. Sharma, G. W. Ding, and M. A. Brubaker, "On the effectiveness of low frequency perturbations," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, ser. IJCAI'19. AAAI Press, 2019, p. 3389–3396.

[17] Z. Jin, J. Zhang, Z. Zhu, X. Wang, Y. Huang, and H. Chen, "Leveraging Information Consistency in Frequency and Spatial Domain for Adversarial Attacks," *arXiv e-prints*, p. arXiv:2408.12670, Aug. 2024.

[18] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," *arXiv e-prints*, p. arXiv:1312.6199, Dec. 2013.

[19] M. Altoub, F. AlQurashi, T. Yigitcanlar, J. M. Corchado, and R. Mehmood, "An ontological knowledge base of poisoning attacks on deep neural networks," *Applied Sciences*, vol. 12, no. 21, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/21/11053

[20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 6000–6010.

[21] A. Aldahdooh, W. Hamidouche, and O. Deforges, "Reveal of Vision Transformers Robustness against Adversarial Attacks," *arXiv e-prints*, p. arXiv:2106.03734, Jun. 2021.

[22] R. Shao, Z. Shi, J. Yi, P.-Y. Chen, and C.-J. Hsieh, "On the Adversarial Robustness of Vision Transformers," *arXiv e-prints*, p. arXiv:2103.15670, Mar. 2021.

[23] S. Paul and P. Chen, "Vision transformers are robust learners," *CoRR*, vol. abs/2105.07581, 2021. [Online]. Available: https://arxiv.org/abs/2105.07581

[24] A. Joshi, G. Jagatap, and C. Hegde, "Adversarial token attacks on vision transformers," *CoRR*, vol. abs/2110.04337, 2021. [Online]. Available: https://arxiv.org/abs/2110.04337

[25] B. Wu, C. Xu, X. Dai, A. Wan, P. Zhang, Z. Yan, M. Tomizuka, J. Gonzalez, K. Keutzer, and P. Vajda, "Visual transformers: Token-based image representation and processing for computer vision," 2020.

[26] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[27] T. Ridnik, E. Ben-Baruch, A. Noy, and L. Zelnik-Manor, "ImageNet-21K Pretraining for the Masses," *arXiv e-prints*, p. arXiv:2104.10972, Apr. 2021.

[28] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv e-prints*, p. arXiv:1412.6980, Dec. 2014.

[29] J. Kiefer and J. Wolfowitz, "Stochastic Estimation of the Maximum of a Regression Function," *The Annals of Mathematical Statistics*, vol. 23, no. 3, pp. 462 – 466, 1952. [Online]. Available: https://doi.org/10.1214/aoms/1177729392

[30] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," *arXiv e-prints*, p. arXiv:1711.05101, Nov. 2017.

[31] J. Yang, K. Z. Luo, J. Li, C. Deng, L. Guibas, D. Krishnan, K. Q. Weinberger, Y. Tian, and Y. Wang, "Denoising Vision Transformers," *arXiv e-prints*, p. arXiv:2401.02957, Jan. 2024.

Backbone Feature Enhancement and Decoder Improvement in HRNet for Semantic Segmentation

HanLei Feng, TieGang Zhong
School of Electronic and Information Engineering, Liaoning Technical University, Huludao Liaoning, 125105, China

***Abstract*—Addressing issues such as the tendency for small-scale objects to be lost, incomplete segmentation of large-scale objects, and overall low segmentation accuracy in existing semantic segmentation models, an improved HRNet network model is proposed. Firstly, by introducing multi-branch deep stripe convolutions, features of multi-scale objects are adaptively extracted using convolutional kernels of different sizes, which not only enhances the model's ability to capture multi-scale objects but also strengthens its perception of the contextual environment.Secondly, to optimize the feature aggregation effect, the axial attention mechanism is adopted to aggregate image features along the x-axis and y-axis directions respectively, effectively capturing long-range dependencies within the global scope, and thus achieving precise positioning of objects of interest in the feature map.Finally, by implementing the progressive fusion-based upsampling strategy, it facilitates the complementary fusion of semantic information and detailed information between adjacent feature maps, thereby enhancing the model's capability to restore fine-grained details in images. Experimental results demonstrate that on the PASCAL VOC2012+SBD dataset, the mean Intersection over Union (mIoU) of the improved HRNet_S model in segmenting lower-resolution images is increased by 1.54% compared to the baseline method. Meanwhile, the improved HRNet_L model achieved a 3.05% increase in mIoU compared to the original model when handling higher-resolution image segmentation tasks on the Cityscapes dataset, and attained the highest segmentation accuracy in 15 out of the 19 different scale classification categories on this dataset.These results indicate that the proposed method not only exhibits high segmentation accuracy but also possesses strong adaptability to multi-scale objects.**

Keywords*—Semantic segmentation; HRNet; multi-branch deep strided convolution; axial attention mechanism; progressive fusion upsampling; multi-scale object adaptability

I. Introduction

Semantic segmentation aims to assign each pixel in an image to a specific semantic category, thereby enabling pixel-level understanding and segmentation of the image. In recent years, with the rapid development of fields such as autonomous driving and medical image processing, semantic segmentation technology has received widespread attention. Semantic segmentation models based on neural networks have become a research hotspot due to their high-precision segmentation capabilities. Although semantic segmentation technology has achieved remarkable application effects in multiple fields, current semantic segmentation methods still face many challenges due to the complex and variable nature of segmentation scenarios and the limitations of the segmentation models themselves. Firstly, the types of objects to be segmented are numerous, and there are significant differences in object scales. Existing semantic segmentation models have difficulties in accurately segmenting object edges and often encounter situations where small-scale objects are ignored or large-scale objects are segmented incompletely. Secondly, most existing models adopt an encoder-decoder architecture. To reduce computational complexity, multiple downsampling operations are performed during the encoder stage, which leads to the loss of a large amount of detailed information that is difficult for the decoder network to effectively recover through learning. Furthermore, existing models have not fully utilized the effective information in feature maps at various levels, making it difficult to strike a balance between shallow detailed information and deep semantic information. In response to the above issues, this paper proposes several improvements based on the HRNet [1] network framework. The main contributions are as follows:

- A Backbone Feature Enhancement Module (BFEM) is designed, which utilizes depth-wise strip convolutions of different scales to adaptively extract features from the backbone network's output. This generates appropriate weights for objects at each scale, addressing the issue of a single convolution kernel's difficulty in adapting to variations in object scales.
- A Flexible Upsampling Mechanism (FUSM) is proposed, which enhances the model's ability to restore image details by facilitating continuous information exchange between adjacent feature maps.
- To overcome the limitations of traditional spatial attention mechanisms in modeling long-distance dependencies, a novel Axial Attention Mechanism (AAM) is proposed. It independently generates attention maps along the length and width dimensions in the spatial dimension and sequentially applies these attention maps to the weighted processing of the input feature maps. While maintaining a lightweight design, it endows the output feature maps with remarkable directional sensitivity, effectively preserving the key positional information in the input features.
- For segmentation tasks involving input images with low resolution, the HRNet backbone network is combined with the Backbone Feature Enhancement Module, the Flexible Upsampling Module, and the Efficient Channel Attention Mechanism to construct the HRNet_S model.Furthermore, to address the challenge of segmenting high-resolution images, the Flexible Upsampling Module is further integrated with the AAM to optimize the decoder network of the HRNet model, resulting in the development of the HRNet_L model.

II. Related Works

To further elevate network performance, current research primarily focuses on three pivotal technological directions: aggregating multi-scale contextual information, enhancing global perception capabilities, and improving detail capture abilities.

In terms of aggregating multi-scale contextual information, multi-scale contextual aggregation networks based on dilated convolutions are widely employed. DeepLabv2 [2] achieves robust segmentation of objects across different scales by introducing atrous spatial pyramid pooling. DeepLabv3 [3] further designs a cascaded module utilizing dilated convolutions, bolstering the network's ability to detect convolutional features across multiple scales. PSPNet [4] harnesses the power of pyramid pooling modules to capture global contextual information for scene parsing tasks. DenseASPP [5] leverages an atrous spatial pyramid structure to extract a broader range of scale feature information. APCNet [6] introduces the Adaptive Pyramid Context Network, constructing multi-scale contextual representations through its meticulously designed Adaptive Context Module (ACM). SCTNet [7] enhances the efficiency of capturing multi-scale contextual information by learning semantic information alignment from Transformer to CNN.R-FPN [8] introduces learnable weights, enabling the network to adaptively utilize effective information from different scale feature maps, thereby improving the effectiveness of small object detection. The improved RCN model proposed by Zhu et al. utilizes the RFN module for multi-scale feature fusion, significantly enhancing the network's ability to recognize microscopic images [9].

To strengthen global perception and capture long-range dependencies, CCNet [10] incorporates a criss-cross attention module, collecting contextual information along criss-cross paths. OCNet [11] proposes an efficient interlaced sparse attention scheme, modeling pixel relationships through sparse relation matrices. DANet [12] adaptively integrates local and global feature dependencies, modeling semantic dependencies in both spatial and channel dimensions. Vision Transformer [13] utilizes a self-attention mechanism, dividing images into patches to construct pixel relationships and accurately capture long-distance dependencies. Building upon this, Swin Transformer [14] and Segformer [15] introduce multi-scale feature extraction and integration methods, enabling models to better understand the dependencies between elements. PIDNet [16] ingeniously utilizes ratio branches to analyze and preserve rich details in high-resolution feature maps, while leveraging integral branches to synthesize local and global contextual information, thereby capturing and processing complex long-range dependencies.

Regarding improving detail capture abilities, DeepLabv3+ [17] introduces early high-resolution feature maps during the decoding stage, enhancing segmentation accuracy at object boundaries. The UNet [18] network enhances high-resolution representations of feature maps by concatenating shallow feature maps with upsampled deep feature maps through skip connections. TransUNet [19] integrates a Transformer architecture into the UNet model, leveraging Transformers to encode image patches into sequences to capture global contextual information.

However, the improvements of most models in the past have often been confined to a specific area, which has somewhat restricted the potential for enhancing their performance [20], [21], [22]. After deeply analyzing the common characteristics of these successful cases, we successfully extracted three core elements for improving model performance. Based on this, we are committed to skillfully integrating these three aspects of improvement in order to achieve a significant leap in model performance. Therefore, we adopted a cascaded HRNet network as the basic building block, aiming to efficiently aggregate multi-scale contextual information. Subsequently, we introduced the Backbone Feature Enhancement Module and a finely designed attention mechanism to further strengthen the model's feature aggregation and representation capabilities. Finally, by introducing the Flexible Upsampling Mechanism, we significantly improved the model's ability to capture details. Compared to improvements in a single aspect, we conducted experiments to deeply explore the interaction mechanisms among different core elements and accordingly achieved an organic integration of various key improvements, thereby making the performance enhancement of the model more significant and comprehensive.

III. Proposed Method

A. Overall Model Architecture

The proposed improved semantic segmentation model based on HRNet is shown in Fig. 1. The network using the upper half of the figure as the model decoder is named HRNet_S, while the network using the lower half of the figure as the model decoder is named HRNet_L.

Firstly, the image is fed into the HRNet feature extraction network to initially obtain the deep semantic information and shallow detail information required by the model. The HRNet feature extraction network is primarily composed of four transition modules and four stage modules. The Transition module is primarily responsible for transforming the number of channels and performing downsampling operations. After processing by this module, the number of feature map channels between adjacent branches differs by a factor of two, and the length and width of the lower-resolution feature maps are half of those of the higher-resolution feature maps. Consequently, the feature maps in each connected Stage exhibit a cascaded pyramid structure. The Stage module is primarily composed of Basic Block modules, whose internal structure is identical to the residual connection modules in the ResNet network, aiming to further extract deeper-level features from the image. Within these modules, continuous skip connections occur between the branches of the cascaded pyramid feature maps to achieve the fusion of low-level and high-level features.

The HRNet_S network enriches the multi-scale characteristics of the lowest and second-lowest resolution feature map branches by introducing the Backbone Feature Enhancement Module (BFEM), which utilizes depth-wise strip convolutions with different kernel sizes. This process extends the semantic information depth of these two branches. Subsequently, the Flexible UpSampling Mechanism (FUSM) upsamples the four feature map branches in ascending order from low to high resolution. During the upsampling process, continuous information exchange occurs between different branches, achieving precise complementarity between semantic information and

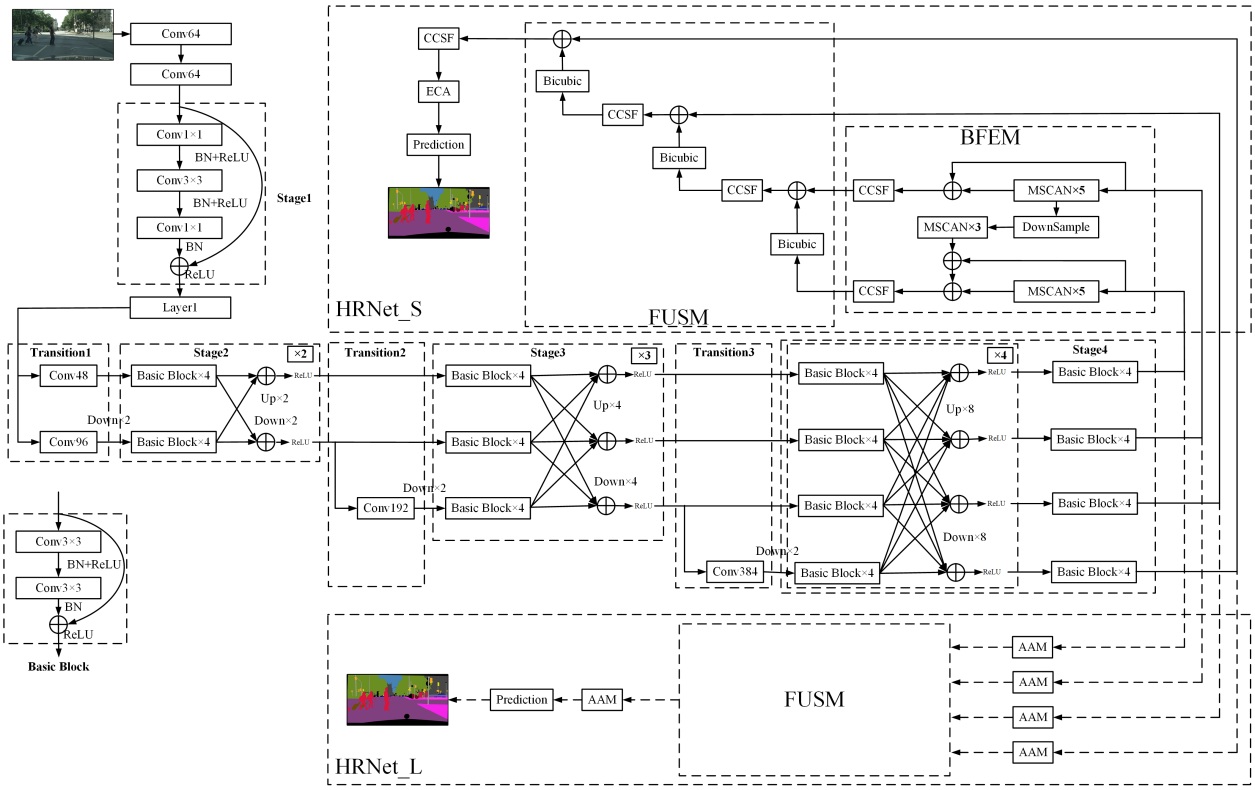


Fig. 1. Improved HRNet network architecture diagram.

detail information. Finally, the feature maps after cross-channel semantic fusion are fed into the Efficient Channel Attention (ECA) module for channel-wise weighting, enabling the network to focus on important features in a targeted manner.

In the HRNet_L network model, the feature maps output by the backbone network undergo feature aggregation along different spatial directions for each channel branch through the Axial Attention Mechanism (AAM), enhancing the model's position sensitivity to objects of interest. Immediately afterward, the feature maps of these four branches undergo upsampling in the same manner as in HRNet_S. Finally, the AAM module is utilized again to further aggregate and enhance the information in the feature maps along the spatial dimension.The HRNet_M model builds upon the HRNet_L model by adding a Cross-Channel Semantic Fusion (CCSF) module after the FUSM, while the remaining parts of the model remain consistent with the HRNet_L model.

B. Backbone Feature Enhancement Module

The structure of the Backbone Feature Enhancement Module (BFEM) is depicted in the BFEM section of Fig. 1. This module utilizes the Multi-Scale Convolutional Attention Network (MSCAN) [23] as the basic unit and strengthens the input features through step-by-step parallel connection. The final output of the sub-low-resolution feature map branch consists of two parts: one is the original feature map of the input branch, and the other is the multi-scale feature map enhanced by the original feature map after passing through five MSCAN modules. For the lowest resolution feature map branch, in addition to performing the same operations as the second-lowest resolution branch, it also concatenates with an output feature map that has undergone a $2\times$ downsampling operation and processing through three MSCAN modules. Since the concatenated branch comes from a deeper layer of the network, it contains richer and more in-depth semantic information, which is crucial for improving the accuracy of the network's category predictions. Meanwhile, the output feature maps from deeper layers have a larger receptive field, enhancing the network's ability to maintain the integrity of object segmentations and recognize large-scale objects. Finally, the semantic information from the two branches is fused through the Cross-Channel Semantic Fusion (CCSF) module. The CCSF module is composed of 1×1 convolution operations, batch normalization operations, and ReLU activation functions, enabling efficient information exchange between channels with a very low number of parameters. This operation can be represented by formula (1).

$$\mathrm{Out} = \mathrm{ReLU}\Big(\mathrm{BN}\Big(\mathrm{Conv}_{1\times1}\Big(F\Big)\Big)\Big) \tag{1}$$

Where, F represents the input feature map; Out represents the output feature map after being processed by the CCSF module.

Fig. 2(a) demonstrates the structure of the Multi-Scale Convolutional Attention (MSCA) module. Firstly, a depthwise convolution with a 5×5 kernel is employed to aggregate local information. Then, depthwise strided convolutions with sizes of 7, 11, and 21 are utilized to capture multi-scale contextual information, aiding the model in understanding contexts across

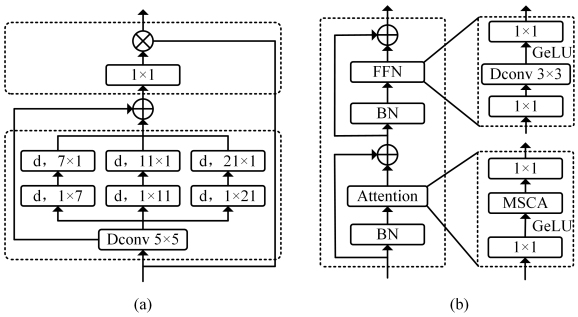


Fig. 2. (a) MSCA module and (b) MSCAN module.

different ranges [23]. Lastly, a 1×1 convolution is applied to model the relationships between different channels in the feature map, and its output is directly used to adjust the weights of the convolutional attention, achieving a weighted reconstruction of the output feature map. Compared to ordinary convolutions, the depthwise strided convolutions utilized in MSCA significantly reduce the number of model parameters. For example, a pair of 1×21 and 21×1 strided convolutions can replace a conventional 21×21 convolution, with the former requiring only 21+21=42 parameters while the latter requires 21×21=441 parameters.Mathematically, MSCA can be concisely described as a combination operation as shown in Eq. (2).

$$\begin{aligned} \text{Att} &= \text{Conv}_{1\times1}\left\{\sum_{i=0}^{3}\text{Scale}_i\Big[\text{DW-Conv}\left(F\right)\Big]\right\} \\ \text{Out} &= \text{Att}\otimes F \end{aligned} \tag{2}$$

Where: F represents the input feature map; Att denotes the attention map; DW-Conv stands for depthwise convolution; Out represents the output weighted attention feature map; $\oplus$ indicates element-wise multiplication operation in matrices; and Scale_i is the i-th branch in the MSCA diagram, where $i \in \{0, 1, 2, 3\}$.

The internal structure of MSCAN, as depicted in Fig. 2(b), primarily comprises three components: Batch Normalization (BN), Feed-Forward Neural Network (FFN), and Multi-Scale Convolutional Attention (MSCA). Among them, the FFN adopts a ResNet-like bottleneck design, aimed at reducing and expanding feature dimensions, thereby minimizing the model's parameter count and computational cost. The attention module shares a similar structure with the FFN module, with the notable difference being the replacement of the original 3×3 depthwise convolution block with the MSCA module to capture richer multi-scale contextual information. The multi-scale features and attention-weighted features output by the attention module are fused through the FFN module, generating the final feature representation.

C. Flexible Upsampling Mechanism

Upsampling operations are crucial steps indispensable for decoder networks to restore image dimensions. Traditional upsampling methods typically involve directly upsampling the deep feature maps or concatenating the upsampled deep feature maps with shallow feature maps. However, this approach has notable limitations: deep feature maps tend to lack detailed information compared to shallow feature maps, and after upsampling, there are significant differences in the distribution of semantic information between the two. This disparity and lack of information often lead to a decrease in final segmentation accuracy. To more effectively utilize the semantic and detailed information from feature maps at different scales and enhance the decoder network's ability to restore detailed information, this paper proposes a bottom-up adjacent feature map priority fusion upsampling strategy. As shown in the FUSM part of Fig. 1, this mechanism can be described by the following steps:

1) For the feature map output from the lowest resolution branch of the HRNet backbone network, a bicubic interpolation algorithm is employed for upsampling. This algorithm considers the grayscale values and their rates of change for the 16 surrounding pixels around the sampling point, generating an enlarged effect that is closer to high-resolution images, effectively preserving image details and mitigating blurring. Subsequently, the upsampled feature map is concatenated with the feature map from the next lower resolution branch along the channel dimension. Since the resolutions of the two are similar, the differences in the spatial distribution of information are relatively small. This facilitates precise alignment of key features between the feature maps, thereby promoting the retention of detailed information and reducing the introduction of noise. Following this, the concatenated feature map is processed by the CCSF module, which models the feature maps from different branches along the channel dimension, enhancing the expressive ability of important features.
2) The fused feature map is then upsampled using bicubic interpolation to match the size of the next higher resolution branch's feature map, and channel concatenation is performed with it. This process is repeated until all feature maps are upsampled to the size of the highest resolution feature map. This not only supplements the details that might be lost during the upsampling process but also enhances the semantic information in the high-resolution feature maps. Through gradual upsampling and fusion, the network can continuously interact and integrate semantic information with detailed information, thereby improving both detail recovery and category prediction accuracy. Finally, the feature map upsampled to the highest resolution will be concatenated with the feature map from the highest resolution branch, completing the entire flexible upsampling process.

Assuming F_1,F_2,F_3 and F_4 are the feature maps output by the backbone feature extraction network from the lowest to the highest resolution branches, respectively, the entire process can be expressed as Eq. (3).

$$\text{Out} = \text{CCSF}\Big(\text{Up}\Big(\text{CCSF}\Big(\text{Up}(F_1)\oplus F_2\Big)\Big)\oplus F_3\Big)\oplus F_4 \tag{3}$$

Where, $\oplus$ denotes concatenation along the channel dimension; Up represents a 2× bicubic interpolation upsampling

operation.

D. Axial Attention Mechanism

In deep learning, attention mechanisms have become a vital component, primarily comprising spatial attention and channel attention. Spatial attention mechanisms enhance focus on important regions by modeling correlations among positions on the feature map and redistributing weights to form spatial masks. In contrast, channel attention focuses on capturing the internal relationships among different channels and intelligently adjusts the importance weights of each channel's features. However, many current attention mechanisms, while fulfilling their functions, also increase the computational complexity of the model. For instance, although the Squeeze-and-Excitation (SE) [24] network can efficiently adjust weights among channels, it overlooks the significance of positional information, which is particularly crucial in tasks requiring precise spatial localization, such as semantic segmentation. The Convolutional Block Attention Module (CBAM) [25] attempts to introduce positional information in the channel dimension through global pooling, but its approach primarily relies on the multi-channel maximum and average values at each position as weight bases.This approach often captures only local information and struggles to model long-range dependencies.

To overcome this limitation, this section proposes an Axial Attention Mechanism (AAM). As shown in Fig. 3, to avoid the potential loss of positional information caused by traditional 2D global pooling, AAM decomposes the computation of spatial attention into two separate 1D feature encoding steps, performed along the length and width directions, respectively. The spatial feature vectors aggregated from these two dimensions not only contain information in their respective directions but also implicitly encode spatial position cues, which can be jointly used to efficiently integrate contextual information of spatial coordinates. The 1D pooling calculation processes along the two directions are shown in Eq. (4).

$$\begin{aligned}\text{H\_AvgPool}(F) &= \frac{1}{H}\left[\sum_{0\le j\le H} x(j,1),\cdots,\sum_{0\le j\le H} x(j,W)\right]\\ \text{W\_AvgPool}(F) &= \frac{1}{W}\left[\sum_{0\le i\le W} x(1,i),\cdots,\sum_{0\le i\le W} x(H,i)\right]\end{aligned} \tag{4}$$

Where, H_AvgPool represents the average pooling operation along the height direction, and W_AvgPool represents the average pooling operation along the width direction; H represents the height of the input image, W represents the width of the input image; F represents the input feature map, and $x(i,j)$ represents the pixel value at coordinate (i,j) in F.

Secondly, the two feature vectors rich in specific directional information are further processed to generate two attention maps. Each attention map focuses on its corresponding spatial direction and can effectively capture the long-range dependencies in that direction within the input feature map. In this way, positional information is cleverly encoded and preserved in the generated attention maps. Finally, AAM applies these two attention maps to the original input feature map through element-wise multiplication at corresponding positions, achieving the reweighting of the feature map. This operation process can be described by Eq. (5).

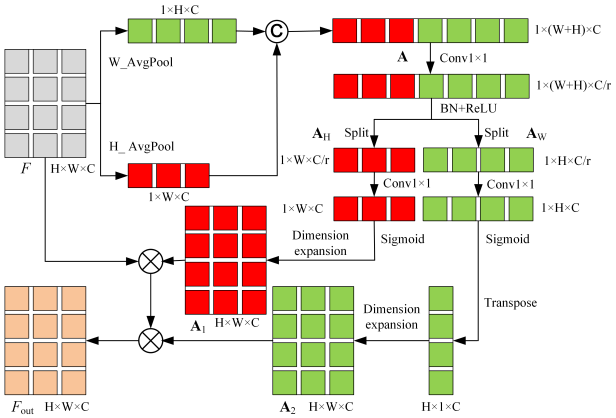


Fig. 3. Axial attention mechanism.

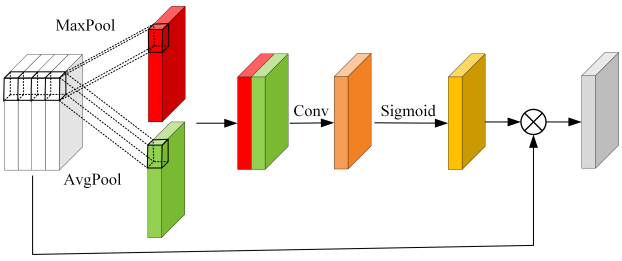


Fig. 4. Traditional spatial attention mechanism.

$$\begin{aligned}&\mathbf{A} = [\text{H\_AvgPool}(F);\text{W\_AvgPool}(F)]\\ &[\mathbf{A}_\text{H};\mathbf{A}_\text{W}]=\text{Split}\Big(\text{ReLU}\Big(\text{BN}\Big(f^{1\times1}(\mathbf{A})\Big)\Big)\Big)\\ &[\mathbf{A}_1;\mathbf{A}_2]=\text{Extend}\left[\sigma\Big(f_1^{1\times1}\Big(\mathbf{A}_\text{H}\Big)\Big);\sigma\Big(f_2^{1\times1}\Big(\mathbf{A}_\text{W}\Big)\Big)'\right]\\ &F_\text{out} = F\otimes\mathbf{A}_1\otimes\mathbf{A}_2\end{aligned} \tag{5}$$

Where, $\mathbf{A}$ represents the feature vector formed by concatenating the features from both the height and width dimensions of the input image;$\mathbf{A}_\text{H}$ and $\mathbf{A}_\text{W}$ represent the intermediate vectors obtained after performing channel dimension reduction on the features from the height and width dimensions, respectively.$\mathbf{A}_1$ and $\mathbf{A}_2$ represent the length and width directional attention matrices obtained by expanding the channel dimension and spatially expanding (replicating the single-dimensional vectors along a certain direction to match the matrix size of F) from $\mathbf{A}_\text{H}$ and $\mathbf{A}_\text{W}$, respectively.$\otimes$ denotes element-wise multiplication of corresponding positions in the matrices.

AAM not only enhances the directional sensitivity of the feature map but also retains crucial positional information, ensuring that the output results more accurately map the target regions in the image. Consequently, when performing dense prediction tasks such as semantic segmentation, AAM can significantly improve the localization accuracy of the model.

Assuming the input feature map has C channels, a height of H, and a width of W, and all convolution operations

employ 1×1 convolutional kernels. In the channel compression stage, AAM uses a compression factor r. Then, the computational complexity of the AAM attention mechanism and the traditional attention mechanism can be estimated using Eq. (6) and (7), respectively, both of which remain at the order of $O(C \cdot H \cdot W)$. Compared to traditional spatial attention mechanisms, AAM effectively captures long-range dependencies while hardly increasing computational complexity. This significant advantage makes AAM more competitive in scenarios involving high-resolution images or high real-time requirements.

$$O(\text{AAM}) = O(4 \cdot C \cdot H \cdot W) + O\Big(2 \cdot \frac{C}{\text{r}} \cdot (H+W)\Big) + O\Big(C \cdot (H+W)\Big) = O(C \cdot H \cdot W) \tag{6}$$

The computational complexity of pooling operations in both the height and width directions is $O\left(2 \cdot C \cdot H \cdot W\right)$; the computational complexity of channel compression and recovery operations is $O\left(2 \cdot \frac{C}{\text{r}} \cdot (H+W)\right)$; the complexity of the Sigmoid function can be considered as $O\left(C \cdot (H+W)\right)$; and the complexity of the two element-wise multiplications can be considered as $O\left(2 \cdot C \cdot H \cdot W\right)$. In practical applications, $\frac{C}{\text{r}}$ can be regarded as a constant, therefore, the above results can be simplified to $O\left(C \cdot H \cdot W\right)$.

The schematic diagram of the traditional spatial attention mechanism is shown in Fig. 4, and its computational complexity can be expressed as Eq. (7).

$$O\left(\cdot\right) = O\left(6 \cdot C \cdot H \cdot W\right) = O\left(C \cdot H \cdot W\right) \tag{7}$$

The computational complexity of max pooling and average pooling is $O\left(2 \cdot C \cdot H \cdot W\right)$; the computational complexity of convolution operations is $O\left(2 \cdot C \cdot H \cdot W\right)$; the computational complexity of the Sigmoid operation is $O\left(C \cdot H \cdot W\right)$; The computational complexity of element-wise multiplication operation can be considered as $O\left(C \cdot H \cdot W\right)$.

IV. Experiment

A. Experimental Dataset

The Cityscapes dataset, jointly provided by three German institutions including Daimler AG, comprises stereo vision data from over 50 cities, featuring a total of 5000 finely annotated and 20000 coarsely annotated images of urban street scenes. In the experiments, only the 5000 finely annotated images are utilized, with 2975 for training, 500 for validation, and 1525 for testing. Each image has a resolution of 1024 pixels × 2048 pixels and is densely annotated with 19 object categories. Since the labels for the test set in the Cityscapes dataset are not publicly available, this paper evaluates the models on the validation set.

The PASCAL VOC2012+SBD dataset is an extension of the PASCAL VOC2012 dataset, obtained by merging PASCAL VOC2012 with the SBD dataset. After removing duplicate images from both datasets, this augmented dataset comprises a total of 12031 annotated images covering 20 different object categories. The images are randomly split into training and validation sets at a 9:1 ratio, with 10827 images for training and 1204 images for validation. Similarly, the models are evaluated on the validation set.

TABLE I. Experimental Hardware and Software Environment

Project Detail	Specification
CPU	16 vCPU Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz
GPU	NVIDIA GeForce RTX 3090 GPU(24GB) × 1
RAM	42GB
CUDA	12.2
Python type	3.8.10
Operating system	Ubuntu 18.04.5 LTS
Development framework	Torch 2.2.1

B. Experimental Environment and Parameter Settings

All networks in the experiment were implemented based on the PyTorch framework, and the hardware and software environments used during the training process are detailed in Table I. The hyperparameter settings during the training process are shown in Table II. The experiments used stochastic gradient descent with momentum for gradient updates. The learning rate decay adopted a cosine annealing schedule, with the specific calculation process described in Eq. (8).

$$lr = min\_lr + 0.5 \cdot (initial\_lr - min\_lr) \cdot \left(1 + \cos\left(\pi \cdot \frac{iter}{total\_iter}\right)\right) \tag{8}$$

Where, $initial\_lr$ represents the initial learning rate; min_lr is the minimum learning rate, which is set to 1/100 of $initial\_lr$; lr represents the current learning rate; $iter$ is the current iteration number; and $total\_iter$ represents the total number of iterations.

For models trained on the Cityscapes dataset, a combination of Focal Loss and Dice Loss was adopted as the loss function. Focal Loss can effectively address the class imbalance issue present in the Cityscapes dataset, while Dice Loss enhances the accuracy of edge prediction. Combining these two losses optimizes the training effect of the model. On the PASCAL VOC2012+SBD dataset, where the class distribution is relatively balanced, a simple CE Loss (Cross-Entropy Loss) was chosen as the loss function.To optimize the training process, bilinear interpolation was used to uniformly resize the image resolution in the Cityscapes dataset to 512 pixels × 1024 pixels, while the image resolution in the PASCAL VOC2012+SBD dataset was adjusted to 512 pixels × 512 pixels. For training all models, a transfer learning strategy was employed, where pre-trained weights on the ImageNet dataset were loaded into the backbone network, and a frozen training phase of 50 epochs was first conducted. For models trained on the PASCAL VOC2012+SBD dataset, the number of training epochs was set to 200, while for models trained on the Cityscapes dataset, the number of training epochs was set to 300.

TABLE II. HYPERPARAMETER SETTINGS

Hyperparameter	PASCAL VOC2012+SBD	Cityscapes
input_pixel	512×512	512×1024
optimizer	SGD	SGD
initial_lr	0.004	0.004
lr_decay_strategy	cos	cos
momentum	0.9	0.9
freeze_iters	50	50
unfreeze_iters	250	150
total_iters	300	200
loss_function	CE Loss	Focal Loss + Dice Loss

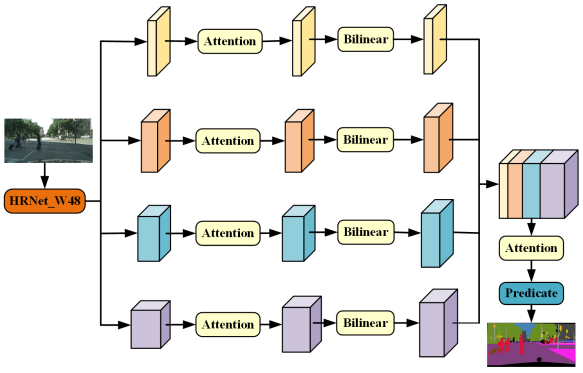


Fig. 5. Schematic diagram of attention mechanism embedding positions.

C. Experimental Results and Analysis

1) Attention selection experiment: To investigate the impact of different attention mechanisms on the segmentation performance of the model, a series of attention mechanisms were sequentially embedded into the "Attention" positions in Fig. 5 for training. These mechanisms include Strip Pooling (SP) [26], CBAM, AAM, SE, and ECA. Analyzing the experimental data in Fig. 6, it can be seen that when the test image resolution is 512 pixels×1024 pixels, the combination of the HRNet network with ECA achieves the best segmentation accuracy, with a 1.12% improvement in mean Intersection over Union (mIoU) compared to the original HRNet model. When the test image resolution is 1024 pixels×2048 pixels, the combination of HRNet with the AAM attention mechanism performs particularly well, with a 2.34% increase in mIoU compared to the original model.

ECA, AAM and CBAM, all have a positive impact on improving the segmentation accuracy of HRNet. Preliminary analysis indicates that when processing lower-resolution test images, channel attention mechanisms can significantly enhance the segmentation accuracy of the model. However, as the resolution of the test images increases, spatial attention mechanisms gradually demonstrate their unique advantage in improving model accuracy. The reason may be that when processing high-resolution images, channel attention mechanisms require deeper spatial information compression, leading to significant loss of spatial pixel information in the feature maps. Due to this information loss, channel attention weights cannot effectively measure the importance of information in the feature maps, resulting in suboptimal weighting effects on the feature maps. However, such issues do not exist when using spatial attention mechanisms.

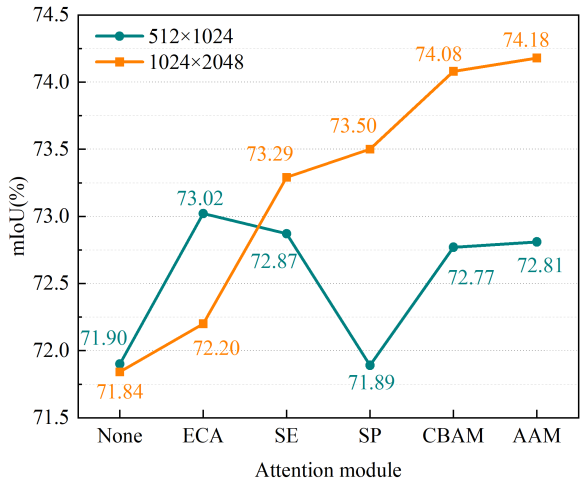


Fig. 6. Comparison of mIoU using HRNet with different attention mechanisms.

TABLE III. ABLATION EXPERIMENT RESULTS ON CITYSCAPES DATASET

BFEM	FUSM	CCSF	ECA	AAM	mIoU *(%)	mIoU ^(%)	Params /M	FLOPs */G	FLOPs ^/T
					72.34	71.79	**65.860**	**376.184**	**1.505**
✓					72.46	72.67	84.342	402.918	1.612
			✓		73.28	73.35	**65.860**	376.232	**1.505**
				✓	73.11	74.62	66.128	376.291	**1.505**
	✓				73.25	72.23	66.646	384.996	1.540
	✓	✓			73.47	73.29	67.166	419.158	1.677
	✓	✓	✓		73.70	74.37	67.166	419.205	1.677
✓	✓	✓	✓		**73.75**	73.93	85.647	445.939	1.784
	✓			✓	73.37	**74.84**	66.914	385.102	1.540
	✓	✓		✓	**73.75**	74.49	67.434	419.265	1.677

[1] "*" indicates that the image resolution used during testing is 512 pixels×1024 pixels, "^" indicates that the image resolution used during testing is 1024 pixels×2048 pixels.
[2] FLOPs stands for Floating-Point Operations, referring to the total number of basic arithmetic operations (addition, subtraction, multiplication, division, etc.) required by the model during execution, which directly reflects the computational complexity of the model.
[3] G represents billion (10^9) level of operations, and T represents trillion (10^{12}) level of operations.

2) Ablation study: To validate the effectiveness of the added modules, this section sequentially adds different (combinations of) modules to the HRNet and conducts experimental verifications. The experimental results are detailed in Table III. To more intuitively demonstrate the improvement effects of each module (and their combinations) on segmentation accuracy (mIoU), the incremental data is plotted into a bar chart as shown in Fig. 7.

When the resolution of the test images is 512 pixels×1024 pixels, the impact of each module on model performance:

FUSM: After introducing this module, the model's mIoU metric improves by 0.91%, while only introducing a small increase in parameters (0.786M) and computations (8.812G). This indicates that FUSM enhances segmentation accuracy while maintaining model efficiency, particularly excelling in detail recovery.

CCSF: Adding this module on top of FUSM further boosts

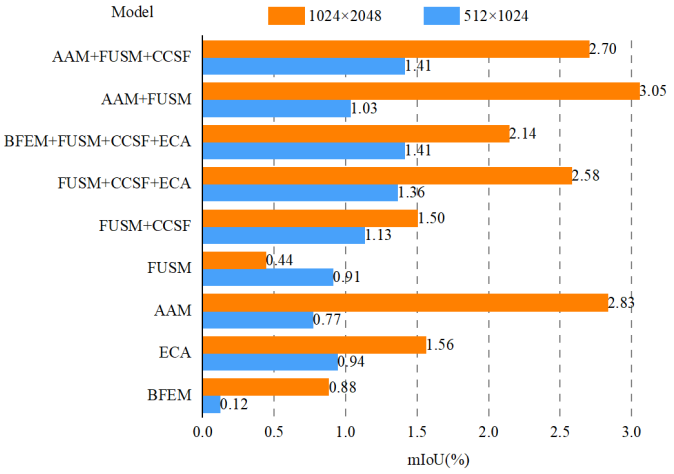


Fig. 7. The improvement in mIoU after integrating various (combinations of) modules into HRNet.

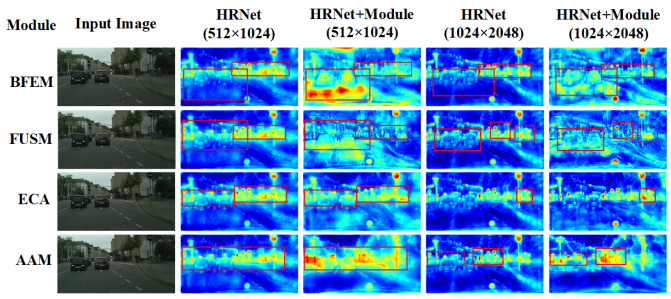


Fig. 8. Heatmap outputs before the fully connected classification layer when combining HRNet with various modules.

the mIoU metric by 0.22%. However, it significantly increases the computational cost by 34.162G. This suggests that while CCSF can improve accuracy, it comes with a relatively high computational overhead.

ECA: Incorporating this module enhances the mIoU metric by 0.27% with virtually no additional parameters or computations, demonstrating ECA's efficiency in capturing crucial information.

BFEM: Upon introduction, the mIoU metric surges to a peak of 73.75%, showcasing the module's remarkable ability to enrich multi-scale contextual information and deepen semantic understanding. However, this substantial improvement is accompanied by a notable increase in both spatial and temporal complexity of the model (an increase of 18.481M parameters and 26.734G total floating-point operations), indicating that higher computational costs are necessary to achieve this level of performance enhancement.

When the resolution of the test images is 1024 pixels$\times$2048 pixels, the impact of each module on model performance:

AAM: After introducing AAM, the mIoU significantly improves by 2.83% with only a small increase in parameters (0.268M) and almost no increase in computational complexity, demonstrating its efficient capability in processing spatial information.

FUSM: Further incorporating FUSM on top of AAM elevates the mIoU to 78.48%, achieving the optimal value. The increases in parameters and computations are relatively small, at 0.786M and 0.035T respectively, indicating that FUSM can still effectively enhance accuracy while maintaining its lightweight advantage at high resolutions.

CCSF: Adding the CCSF module to the AAM+FUSM combination unexpectedly decreases the mIoU to 74.49%, underperforming expectations, especially in high-resolution scenarios.Possible reasons for this phenomenon include: increased information redundancy or conflicts between feature maps, a high computational burden due to elevated model complexity, and complex interactions between modules that affect the model optimization process.

Fig. 8 shows heatmaps output by HRNet and its combinations with different modules, revealing the impact of each module on changes in target attention. The original HRNet heatmap primarily focuses on small targets such as bicycles and pedestrians, indicating its advantage in recognizing small-scale objects. After introducing the BFEM module, the heatmap not only continues to pay attention to small-scale targets but also starts to focus on large targets like cars and roads, suggesting that BFEM enhances the model's recognition capabilities for multi-scale targets. The addition of the FUSM module makes the temperature distribution of the heatmap more uniform and emphasizes object edges more strongly, indicating that FUSM improves the model's scene and contour perception abilities. The ECA module enables the model to more precisely focus on critical targets, contributing to enhanced feature extraction capabilities. Lastly, the AAM module clarifies the location information of objects of interest by enhancing the contrast between hot and cold regions, thereby improving the model's segmentation accuracy.

In summary, each module exhibits unique characteristics in image segmentation tasks, and their effects are significantly influenced by resolution and application scenarios. FUSM and ECA demonstrate a good balance between precision improvement and complexity control across different resolutions. While CCSF and BFEM can enhance performance, the associated increase in computational complexity cannot be overlooked. AAM is particularly efficient at high resolutions, highlighting its advantage in spatial information processing. In practical applications, the optimal combination of modules should be flexibly selected and configured based on specific requirements and environmental conditions to achieve the best balance between performance and computational efficiency.

3) Comparison with State-of-the-Art methods: To validate the performance of the proposed method in semantic segmentation, this section compares it with various state-of-the-art approaches on two standard datasets: Cityscapes and PASCAL VOC2012+SBD. The compared methods include DeepLabv3_R (based on ResNet50), DeepLabv3_X (based on Xception), PSPNet, SegFormer, SeaFormer-S, SeaFormer-B, SCTNet-S, and HRNet. The segmentation results on Cityscapes and PASCAL VOC2012+SBD datasets are presented in Tables IV and V, respectively.

Analyzing the data in Table IV, it can be observed that at a resolution of 512 pixels$\times$1024 pixels, both HRNet_S and HRNet_M achieve an optimal segmentation accuracy (mIoU)

TABLE IV. COMPARISON OF THE IMPROVED MODEL WITH STATE-OF-THE-ART METHODS ON THE CITYSCAPES DATASET

Model	mIoU *(%)	mIoU ˆ(%)	Params /M	FLOPs */G	FLOPs ˆ/T
DeepLabv3+_R[17]	70.36	71.33	**40.354**	732.532	2.930
PSPNet[4]	68.57	69.22	48.957	749.149	2.996
SegFormer[15]	72.05	72.32	47.238	**286.349**	1.145
SeaFormer-S[27]	-	70.70	-	-	**0.02**
SeaFormer-B[27]	-	72.70	-	-	0.03
SCTNet-S[7]	72.80	-	4.6	-	-
HRNet[1]	72.34	71.79	65.860	376.184	1.505
HRNet_S(ours)	**73.75**	73.93	85.647	445.939	1.784
HRNet_M(ours)	**73.75**	74.49	67.434	419.265	1.677
HRNet_L(ours)	73.37	**74.84**	66.914	385.102	1.540

TABLE V. COMPARISON OF THE IMPROVED MODEL WITH STATE-OF-THE-ART METHODS ON THE PASCAL VOC2012+SBD DATASET

Model	mIoU(%)	Params/M	FLOPs/G
DeepLabv3+_R[17]	75.68	**40.354**	366.275
DeepLabv3+_X[17]	77.82	54.714	486.773
PSPNet[4]	75.45	48.958	374.634
SegFormer[15]	**80.30**	47.239	**143.199**
HRNet[1]	77.42	65.861	188.116
HRNet_S(ours)	78.93	85.648	222.993
HRNet_M(ours)	77.62	67.435	209.674
HRNet_L(ours)	76.49	66.915	192.529

of 73.75%, surpassing the novel SegFormer model by 1.70%. Even when compared with the recent SCTNet-S model, it remains 0.95% higher. When processing images with a higher resolution of 1024 pixels×2048 pixels, the mIoU of the HRNet_L model climbs to 74.84%, outperforming SegFormer by 2.52%. Even when compared with the recent SeaFormer-B model, it remains 2.14% higher. However, in terms of parameter count and computational complexity, HRNet shows slight increases compared to SegFormer and does not have an advantage over SegFormer-B in terms of computational efficiency. Therefore, lightweight design of the model is a key focus for future research.

TABLE VI. WHEN THE TEST IMAGE RESOLUTION IS 1024 PIXELS×2048 PIXELS, THE PER-CATEGORY MIOU OF DIFFERENT MODELS ON THE CITYSCAPES DATASET (%)

Category	PSPNet	DeepLabv3+_R	SegFormer	HRNet	HRNet_L
road	91.31	92.74	92.89	93.02	**93.23**
sidewalk	72.19	71.46	73.25	72.76	**75.30**
building	86.71	87.80	87.88	88.80	**88.89**
fence	41.60	44.65	50.08	**54.12**	48.70
pole	51.69	52.68	51.79	**56.63**	56.26
traffic light	53.55	59.22	56.51	60.85	**62.84**
traffic sign	62.40	65.93	66.04	66.95	**68.78**
vegetation	66.43	70.09	70.39	71.63	**72.79**
terrain	89.55	90.01	90.38	90.52	**90.67**
sky	57.08	56.20	57.27	58.53	**59.42**
diningtable	89.00	90.17	90.32	**90.59**	90.41
person	75.28	77.64	77.53	79.15	**79.80**
rider	59.33	63.01	61.00	66.34	**68.13**
car	92.26	92.23	92.17	92.51	**93.58**
truck	55.60	60.73	70.14	63.52	**78.18**
bus	82.32	83.16	81.17	80.07	**85.34**
train	71.06	69.22	77.60	46.66	**81.89**
motorcycle	53.49	62.01	62.54	**64.28**	61.09
bicycle	71.90	73.47	72.00	73.39	**74.37**
background	61.71	64.06	65.27	65.53	**67.12**

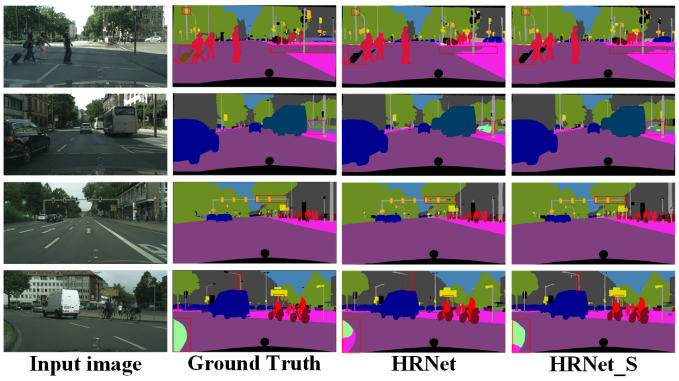


Fig. 9. The segmentation effect images of different models on the Cityscapes dataset when the test image resolution is 512 pixels×1024 pixels.

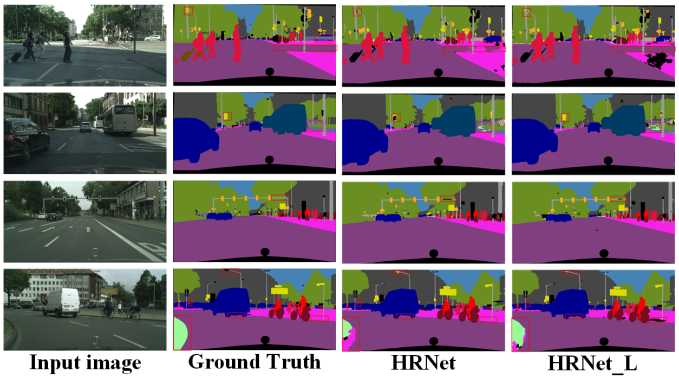


Fig. 10. The segmentation effect images of different models on the Cityscapes dataset when the test image resolution is 1024 pixels×2048 pixels.

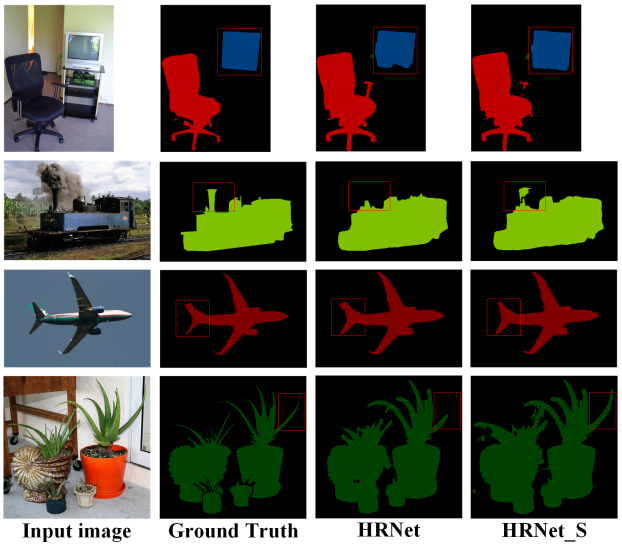


Fig. 11. Segmentation effect images of different models on the PASCAL VOC2012+SBD dataset.

As shown in Table V, on the PASCAL VOC2012+SBD dataset, for low-resolution images of 512 pixels×512 pixels, the mIoU of the HRNet_S model reaches 78.93%. Although

slightly lower in segmentation accuracy compared to the SegFormer model, HRNet_S has significantly narrowed the gap with SegFormer, demonstrating its competitiveness in low-resolution image segmentation tasks. However, from the perspective of algorithmic complexity, there is still room for further optimization of the HRNet_S model.

Combing the experimental data from Tables IV and V, from the perspectives of segmentation accuracy and model complexity, HRNet_M can be considered as a balanced choice between HRNet_S and HRNet_L. In terms of increasing computational complexity, the order is: HRNet_L, HRNet_M, HRNet_S. Additionally, the segmentation accuracy of HRNet_M also falls between that of HRNet_S and HRNet_L across different resolutions.

As shown in Table VI, which presents the category-wise mIoU data of various models on the Cityscapes dataset when tested on images with a resolution of 1024 pixels$\times$2048 pixels, it can be observed that HRNet_L significantly improves the segmentation accuracy for large-scale objects such as trucks, buses, and trains at this high resolution of 1024 pixels$\times$2048 pixels. Furthermore, HRNet_L achieves the highest accuracy in 15 out of the 19 different scale-varying categories.

To intuitively demonstrate the superiority of the proposed method in semantic segmentation tasks, this section comprehensively evaluates its segmentation effects through visual comparison experiments with the baseline method, HRNet. Fig. 9 and 10 showcase the segmentation effects of the models on Cityscapes dataset for test images with different resolutions. From these two figures, it can be observed that HRNet exhibits some shortcomings in image segmentation, particularly in road contours, traffic sign recognition, and segmentation accuracy of elongated objects such as streetlights. Additionally, it tends to confuse vegetation with terrain. However, these deficiencies are significantly improved in the proposed enhanced models, HRNet_S and HRNet_L. Fig. 11 further presents the segmentation effects of the models on the PASCAL VOC2012+SBD dataset. On this dataset, HRNet_S notably enhances the segmentation accuracy of TV and airplane wing boundaries compared to the original HRNet. Simultaneously, it exhibits more refined segmentation capabilities for complex structures such as ship funnels and vegetation leaves. This indicates that the proposed method comprehends contextual semantic information more comprehensively than the original model, validating its effectiveness in improving segmentation accuracy and detail preservation.

Based on the HRNet network, this paper introduces specific modules and successfully develops a model suitable for image segmentation across different resolutions. Specifically, the HRNet_S model demonstrates exceptional segmentation performance when dealing with lower-resolution images. However, due to its relatively high number of parameters and computational complexity, when handling higher-resolution image segmentation tasks, the significant increase in computational load becomes particularly evident, which may lead to inadequate model training and subsequently affect its final performance. Additionally, the Efficient Channel Attention Mechanism employed in HRNet_S may overly compress the valuable information in high-resolution images, adversely impacting the model's ultimate performance. In contrast, the HRNet_L model is more lightweight, thus having an advantage when dealing with higher-resolution image segmentation tasks. Its use of the AAM attention mechanism effectively aggregates valuable information in high-resolution images, which is a key factor contributing to its superior performance. However, compared to the HRNet_S model, the HRNet_L model lacks depth in semantic information, making its advantage less pronounced when handling lower-resolution image segmentation tasks.

Therefore, in practical production and life, we should select the appropriate variant of the HRNet model based on factors such as the resolution of the dataset and the application scenario, in order to achieve the best segmentation results and ensure that the model can perform optimally in image segmentation tasks across different resolutions.

V. Conclusion

Addressing the challenges faced by existing semantic segmentation networks in handling multi-scale objects, such as the tendency to lose small-scale objects, incomplete segmentation of large-scale objects, and low overall segmentation accuracy, this paper proposes an improved method based on the HRNet network. Firstly, a backbone feature enhancement module is introduced using deep stripe convolutions, which enhances the network's adaptability to multi-scale objects, overcomes the limitations of a single convolutional kernel in feature extraction, expands the model's perception range of contextual information, and enhances the network's ability to understand complex scenes. Secondly, an axial attention mechanism is employed to model the global dependency relationships within the feature maps output by the backbone network, enabling precise localization of regions of interest. Lastly, a flexible upsampling mechanism is adopted, leveraging the complementary fusion of semantic and detail information between adjacent feature maps, to effectively restore target detail information in the decoder network. Experimental results show that the proposed algorithm achieves the highest segmentation accuracy compared to other algorithms on the Cityscapes dataset. Similarly, the segmentation accuracy of the proposed algorithm on the PASCAL VOC 2012+SBD dataset is also outstanding, verifying the effectiveness of the proposed method. Further ablation studies also confirm the contributions of each improved component to enhancing the overall performance. The relevant implementation code for this paper has been uploaded to the GitHub platform (link: https://github.com/HanLeiFeng/HRNet_Series.git) for learning and exchange.

References

[1] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5693–5703.

[2] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[3] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[4] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[5] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, "Denseaspp for semantic segmentation in street scenes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3684–3692.

[6] J. He, Z. Deng, L. Zhou, Y. Wang, and Y. Qiao, "Adaptive pyramid context network for semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 7519–7528.

[7] Z. Xu, D. Wu, C. Yu, X. Chu, N. Sang, and C. Gao, "Sctnet: Single-branch cnn with transformer semantic information for real-time segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 6, 2024, pp. 6378–6386.

[8] C. Peng, M. Zhu, H. Ren, and M. Emam, "Small object detection method based on weighted feature fusion and csma attention module," *Electronics*, vol. 11, no. 16, p. 2546, 2022.

[9] M. Zhu, J. Wang, A. Wang, H. Ren, and M. Emam, "Multi-fusion approach for wood microscopic images identification based on deep transfer learning," *Applied Sciences*, vol. 11, no. 16, p. 7639, 2021.

[10] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "Ccnet: Criss-cross attention for semantic segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 603–612.

[11] Y. Yuan, L. Huang, J. Guo, C. Zhang, X. Chen, and J. Wang, "Ocnet: Object context for semantic segmentation," *International Journal of Computer Vision*, vol. 129, no. 8, pp. 2375–2398, 2021.

[12] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3146–3154.

[13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[14] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10 012–10 022.

[15] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," *Advances in neural information processing systems*, vol. 34, pp. 12 077–12 090, 2021.

[16] J. Xu, Z. Xiong, and S. P. Bhattacharyya, "Pidnet: A real-time semantic segmentation network inspired by pid controllers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 19 529–19 539.

[17] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.

[19] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[20] C. Ziwen, K. Patnaik, S. Zhai, A. Wan, Z. Ren, A. Schwing, A. Colburn, and L. Fuxin, "Autofocusformer: Image segmentation off the grid," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 18 227–18 236.

[21] W. Wang, J. Dai, Z. Chen, Z. Huang, Z. Li, X. Zhu, X. Hu, T. Lu, L. Lu, H. Li *et al.*, "Internimage: Exploring large-scale vision foundation models with deformable convolutions," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 14 408–14 419.

[22] A. Hassani and H. Shi, "Dilated neighborhood attention transformer," *arXiv preprint arXiv:2209.15001*, 2022.

[23] M.-H. Guo, C.-Z. Lu, Q. Hou, Z. Liu, M.-M. Cheng, and S.-M. Hu, "Segnext: Rethinking convolutional attention design for semantic segmentation," *Advances in Neural Information Processing Systems*, vol. 35, pp. 1140–1156, 2022.

[24] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[25] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[26] Q. Hou, L. Zhang, M.-M. Cheng, and J. Feng, "Strip pooling: Rethinking spatial pooling for scene parsing," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 4003–4012.

[27] Q. Wan, Z. Huang, J. Lu, G. Yu, and L. Zhang, "Seaformer: Squeeze-enhanced axial transformer for mobile semantic segmentation," *arXiv preprint arXiv:2301.13156*, 2023.

Machine Learning Approach to Identify Promising Mountain Hiking Destinations Using GIS and Remote Sensing

Lahbib Naimi[1], Charaf Ouaddi[2], Lamya Benaddi[3], El Mahi Bouziane[4], Abdeslam Jakimi[5], Mohamed Manaouch[6]
Software Engineering and Information Systems Engineering Team, Department of Informatics,
Faculty of Sciences and Techniques of Errachidia, Moulay Smail University, Errachidia 52000, Morocco[1,2,3,4,5]
Department of Geography, Faculty of Humanities and Social Sciences,
Ibn Tofail University, Kenitra 14000, Morocco[6]

***Abstract*—The objective of this study is to address the complex task of identifying optimal locations for mountain hiking sites in the Eastern High Atlas region of Morocco, considering topographical factors. The study assesses the effectiveness of a commonly used machine learning classifier (MLC) in mapping potential mountain hiking areas, which is crucial for promoting and enhancing tourism in the area. To begin with, an extensive inventory of 120 mountain hiking sites was conducted, and precise measurements of three topographical parameters were collected at each site. Subsequently, a machine learning algorithm called Bagging was employed to develop a predictive model. The model achieved a high performance, with an area under the curve (AUC) value of 0.93. The model effectively identified favorable areas, encompassing around 24% of the study region, which were predominantly located in the western part. These areas were characterized by mountainous terrain, shorter slopes, and higher altitudes. The research findings provide valuable guidance to decision-makers, offering a roadmap to enhance the discovery of mountain hiking sites in the region.**

Keywords*—Machine learning; mountain hiking; AI-based tourism; GIS; remote sensing; tourism; bagging algorithm; decision-making

I. Introduction

Mountain hiking, also known as mountain trekking or mountaineering, is an outdoor recreational activity that involves exploring and traversing mountainous regions on foot. It typically involves escalating and descending mountains or traversing steep and rugged terrain. Mountain hiking often requires physical endurance, technical skills, and knowledge of mountaineering techniques such as navigation, rock climbing, and rope handling. Hikers engage in this activity to experience the natural beauty of mountain landscapes, challenge themselves physically and mentally, and achieve personal goals. Mountain hiking can range from leisurely day hikes on well-marked trails to multi-day expeditions that may involve camping and navigating through remote and challenging environments.

For this purpose, mountaineering has been addressed in several studies regarding its relationship with tourism, the environment, and the emotional aspects of mountaineers. The study conducted by Wang et al. [1] focuses on the contribution of mountaineering to the creation of lasting and impactful tourism experiences. By utilizing Mount Huangshan in China as a case study, the research explores how sensory perceptions, physical engagement, and emotional connections play a pivotal role in shaping the memories of tourists who engage in hiking activities within this renowned mountainous region.

On the other hand, the research conducted by Galiakbarov et al. [2] delved into the psychological and spiritual aspects of mountaineering, using calling theory to explore why individuals engage in this activity. Their research sheds light on the deeply personal motivations behind mountaineering as a purposeful and transformative experience. However, the study lacks focus on geographic and environmental factors that influence the suitability of certain regions for mountaineering, creating a gap that this research aims to address.

Several studies have also investigated the environmental aspects of mountaineering. Naseri et al. [3] examined the environmental, social, and economic dimensions of sustainable development in mountaineering tourism. Their work emphasized the need for sustainable practices to protect fragile mountain ecosystems while fostering tourism. However, geographic-specific methodologies for identifying suitable hiking sites, particularly in under-researched areas like the High Atlas of Morocco, remain limited. This gap becomes especially relevant in regions with rich biodiversity and geological formations, where tourism can play a pivotal role in both economic development and environmental conservation.

Liu et al. [4] explored the connection between leisure participation in mountaineering and environmentally responsible behavior, focusing on how emotional bonds with landscapes (place attachments) influence conservation attitudes. This emphasizes the importance of understanding not only the environmental but also the cultural and emotional factors that drive sustainable tourism development.

Similarly, Rogerson et al. [5] traced the historical evolution of mountaineering tourism in South Africa, showing how cultural and historical significance contributes to the development of special interest tourism. While these studies offer valuable insights, they do not address how modern technological tools like machine learning algorithms can assist in identifying new areas for mountain tourism based on geographical and topographical factors.

In Morocco, research on mountain hiking tourism remains scarce, particularly in the eastern High Atlas region. A notable exception is the study by Kchikach et al. [6], which focuses

on the UNESCO M'Goun Geopark in the central High Atlas. The eastern High Atlas, however, despite being known for its jagged peaks, deep gorges, and ancient rock structures, has not been extensively studied in the context of tourism development. This region holds significant potential for mountain hiking and climbing due to its unique geological formations and natural beauty, making it an ideal candidate for tourism-related research.

To address this gap, this study introduces a novel approach by leveraging machine learning algorithms, specifically Bagging, to identify potential hiking sites in the eastern High Atlas. Bagging, a robust ensemble learning method, was selected due to its ability to improve the accuracy and stability of predictions, particularly in geographic modeling tasks where data variability is high.

Furthermore, the study uses the Area Under the Curve (AUC) and Receiver Operating Characteristic (ROC) metrics to evaluate model performance, ensuring a rigorous assessment of the results. These methods are widely accepted for evaluating the accuracy of predictive models, particularly in machine learning-based geographic studies. The combination of AUC and ROC analysis with Bagging ensures a comprehensive approach to identifying suitable mountain hiking locations based on geographic data.

In addition to the technical aspects, this research has broader implications for sustainable tourism development in Morocco. By identifying suitable hiking locations, the study contributes to the sustainable management of natural resources and helps promote ecotourism in underdeveloped areas. Sustainable tourism practices, which are essential for preserving the environmental and cultural heritage of the region, can be informed by the findings of this study. As tourism continues to grow in the High Atlas, ensuring that it is developed in an environmentally responsible manner is crucial for maintaining the long-term viability of these destinations.

The implications of this research extend beyond the immediate identification of suitable hiking sites. By integrating advanced machine learning techniques into the tourism development process, this study establishes a precedent for leveraging technology to advance the tourism sector in Morocco and globally [7], [8], while also enhancing environmental conservation efforts.

The findings will not only support sustainable tourism initiatives but also contribute to the long-term preservation of the High Atlas's unique ecosystems and cultural heritage. This alignment of tourism with conservation principles can foster a more responsible tourism industry, encouraging both local communities and policymakers to prioritize sustainable practices that benefit the environment and the economy

II. Materials and Methods

A. Study Area

The upper Ziz area in the eastern High Atlas region of southeastern Morocco (Fig. 1) experiences a semi-arid climate with cold winters and dry summers, as indicated by Manaouch et al. [9]. The average annual precipitation ranges from 119 to 377 mm, and temperatures vary between 10.2 and 19.2°C, as documented by Manaouch et al. [10]. The elevation ranges from 1023 to 3687 meters above sea level, with slopes ranging from 0 to 66°C, as noted by Manaouch et al. [11].

This region boasts numerous mountain hiking sites that showcase its unique geological and geomorphological features. These sites provide opportunities for tourism, offering insights into the region's geological history. Among these sites, the Gorges of Ziz are particularly remarkable. Carved by the Ziz river over millennia, these limestone gorges feature towering cliffs and winding canyons, creating a breathtaking landscape. Their geological significance and picturesque beauty make them a magnet for geologists and nature enthusiasts alike.

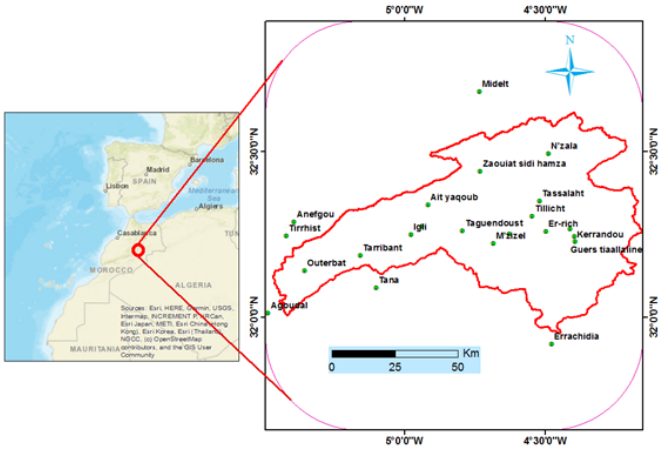


Fig. 1. Location of the eastern high atlas of upper Ziz in Morocco.

B. Data Used and Mountain Hiking Sites Inventory

Accurate identification and selection of mountain hiking sites require a thorough analysis of the phenomenon and its underlying factors [12]. This study conducted extensive field visits in the eastern High Atlas of upper Ziz, SE Morocco, identifying and modeling a total of 120 mountain hiking sites. An equal number of non-mountain hiking sites were also considered to ensure a balanced approach (Fig. 2).

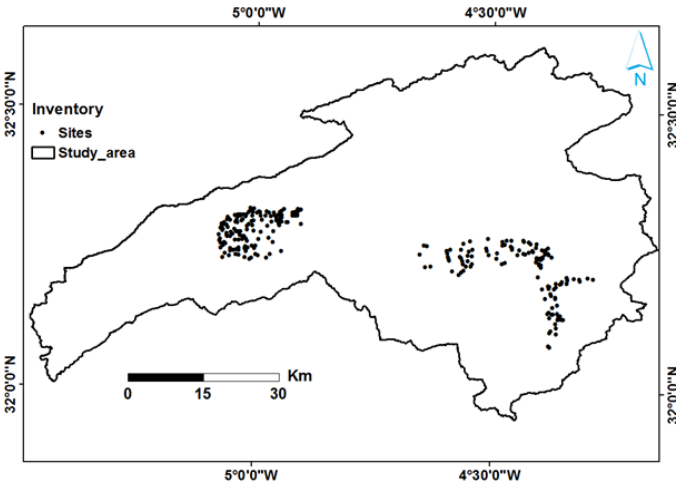


Fig. 2. Mountain hiking sites inventory map.

The study incorporated three conditioning factors (slope, length of slope, and elevation) as independent variables influencing the suitability of areas for mountain hiking sites. These variables were measured for the 240 selected sites. The dataset was divided into training data (70%) and validation data (30%) following established spatial modeling conventions.

Data processing was performed using ArcGIS 10.5 and SPSS Statistics 26 software, which enabled the transformation of geospatial data into a tabular format. The dataset was prepared for analysis by importing data, defining attributes, executing geoprocessing tools, configuring parameters, and performing geoprocessing operations. Model validation was conducted using the Area Under the Curve (AUC) of the receiver operating characteristic (ROC), a common method in geospatial modeling studies to assess and validate model reliability [10].

C. Mountain Hiking Sites (MHS)

Mountain hiking sites (MHS) are of paramount importance in various fields such as environmental education, recreation, and sports [13]. These sites not only offer opportunities for exploration and the appreciation of natural landscapes but also serve as key locations for ecological studies, conservation efforts, and physical fitness activities [14]. Canyons, cliffs, and other geological formations are significant not just for their scenic beauty but also for the insight they provide into the Earth's geological history, making them valuable for scientific research and outdoor education [15].

In addition to their value in education and conservation, MHS contribute to local economies by supporting activities such as mountaineering, hiking, and outdoor excursions, which often require the development of infrastructure such as trails, shelters, and conservation measures. These efforts foster sustainable practices and environmental stewardship while providing recreational opportunities [16].

Fig. 3. Mountainous hiking sites and mountain hikers encountered in the upper Ziz.

While mountain hiking sites are frequently associated with tourism, their broader significance spans disciplines like environmental science, geography, and physical education [17]. In particular, these sites are ideal for studying geomorphological processes, fostering a deeper understanding of natural erosion and sedimentation patterns [18]. Mountain hiking activities, especially in regions like the upper Ziz valley, support not only recreation but also health and well-being through physical fitness and mental restoration [19].

Fig. 3 provides visual examples of mountain hiking sites and hikers along the Ziz valley, showcasing their versatility and value across multiple domains.

III. Methodology

Fig. 4 depicts a flowchart that outlines the sequential procedure employed to evaluate the suitability of mountain hiking locations utilizing an advanced machine learning algorithm. The methodology begins with gathering data from diverse sources, including Google Earth images, field surveys, geological maps, Landsat 8 OLI images, and previous studies. By consolidating this information, a comprehensive database is created, serving as the foundation for the analysis.

Key mountain hiking conditioning factors (MHCFs) such as slope, length of slope, and elevation are identified and incorporated into the study. These factors are crucial in determining the suitability of hiking locations. An inventory map is developed, detailing 240 sites that serve as reference points for the evaluation process. The MHCFs undergo geoprocessing, which involves manipulating spatial data to derive meaningful insights. This step ensures that the data is in a suitable format for further analysis. The processed data is stored in a centralized database, facilitating convenient access and management.

To build and validate the machine learning model, the data is split into training and testing sets, with 70% allocated for training and 30% for testing. This division is essential for developing a robust and reliable model. The process also includes calculating frequency ratios, normalizing data, examining spatial relationships, and preparing data layers, all of which ensure the data is ready for analysis.

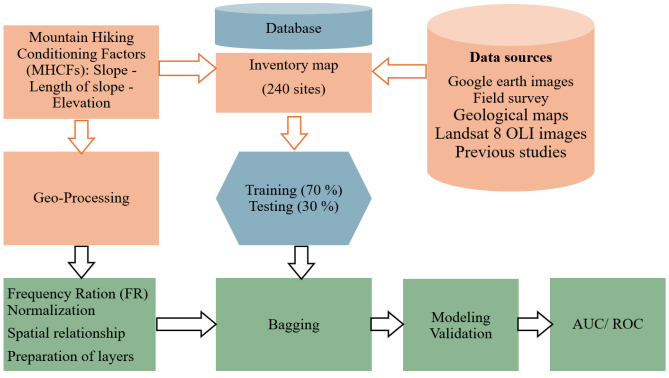


Fig. 4. Flowchart of the modeling strategy used.

The final step involves evaluating the model's performance using AUC (Area Under the Curve) and ROC (Receiver Operating Characteristic) metrics. These metrics provide a measure of the model's accuracy and effectiveness, ensuring the methodology yields reliable results for assessing mountain hiking locations. The subsequent sections offer a comprehensive and detailed explanation of this methodology, providing a deeper understanding of the approach.

A. Mountain Hiking Conditioning Factors (MHCFs)

Topographic factors such as slope, slope length, and elevation significantly impact the suitability of mountainous areas for hiking. These factors influence the characteristics and spatial distribution of land formations. The following sections discuss each factor in detail and provide visual maps representing them.

1) Elevation: Elevation is a crucial attribute in mountain hiking site identification as it helps differentiate various landforms. Mountain hiking sites may vary significantly in elevation, influencing the climate, vegetation, and geological processes of the area. Surveying techniques, satellite-based systems (e.g. GPS), and remote sensing methods like LiDAR or radar are commonly used to measure elevation. Fig. 5 shows the elevation map of the study area.

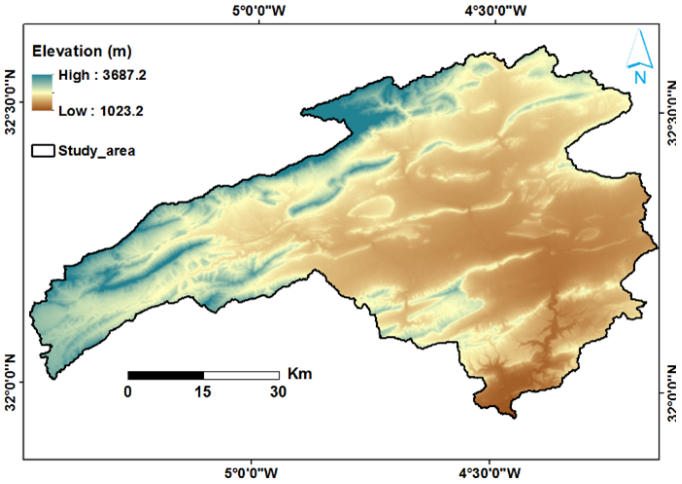


Fig. 5. MHCF map: Elevation.

2) Slope: Slope refers to the steepness or inclination of a landform's surface, a key factor in evaluating hiking site suitability. The slope is calculated in ArcGIS using the variation in elevation between adjacent cells within a Digital Elevation Model (DEM) raster dataset. The formula for calculating slope is:

$$\text{Slope } (\%) = \left(\frac{\text{rise}}{\text{run}}\right) \times 100 \tag{1}$$

Where:

- *rise* represents the change in elevation between cells.
- *run* represents the horizontal distance between the cells.

Fig. 6 illustrates the slope map derived from the DEM.

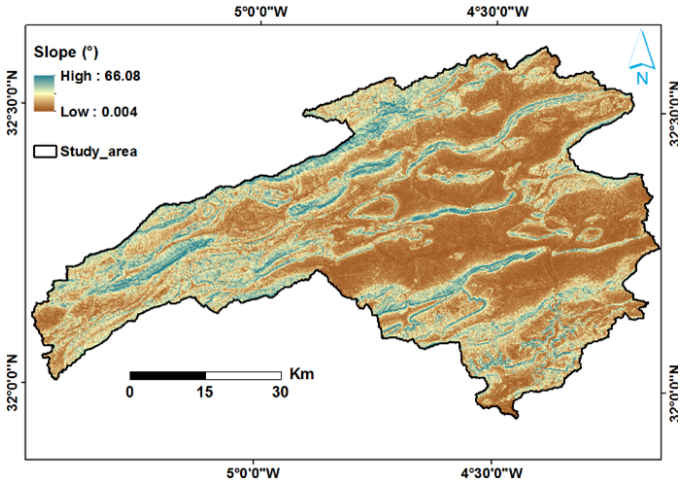


Fig. 6. MHCF map: Slope.

3) Length of Slope (LS): The length of the slope (LS) helps characterize landforms within mountain hiking sites. LS is derived from the DEM using ArcGIS 10.5 software based on the methodology developed by Bizuwerk et al. [20], following Wischmeier and Smith's formula:

$$\text{LS} = \left(\frac{L}{22.13}\right)^{m} \times \left(65.41 \sin^2(S) + 4.56 \sin(S) + 0.065\right) \tag{2}$$

where,

- L is the slope length.
- S is the slope gradient (expressed as a percentage).
- m is a constant value that varies depending on the range of slope gradients.

The resulting LS map is shown in Fig. 7.

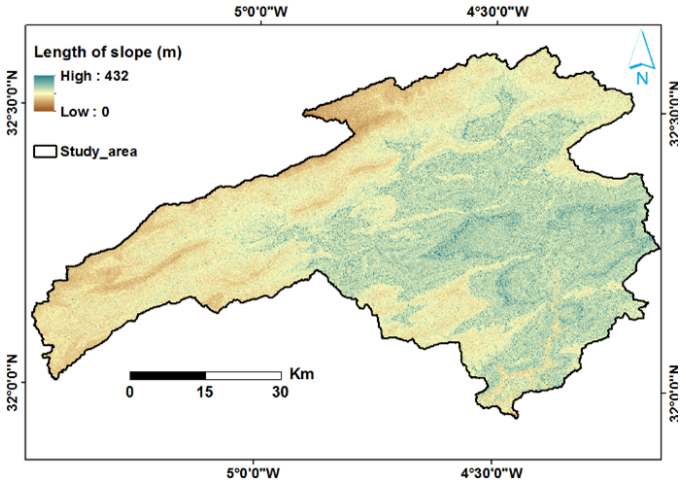


Fig. 7. MHCF map: Length of slope.

B. Machine Learning Algorithm and Frequency Ratio (FR)

1) Bagging algorithm: The Bagging algorithm (Bootstrap Aggregating) comprises two main steps, as described by Breiman et al. [21]. In the first step, multiple training datasets are created by resampling from the original data, a process known as bootstrapping. This involves randomly selecting observations from the original dataset with replacement, allowing some observations to be repeated while others may be omitted. This resampling technique helps to reduce variance and improve the stability of the model.

In the second step, multiple models are constructed using these bootstrapped training datasets. Each model is trained independently, allowing them to capture different aspects of the data. The final predictions are then obtained by aggregating the results of these individual models, typically through averaging for regression tasks or majority voting for classification tasks. This aggregation process tends to produce a more robust and accurate model than any single constituent model.

Bagging has been widely applied in geospatial modeling tasks, where the inherent variability in spatial data can significantly impact model performance. For instance, studies like [22] and [23] have demonstrated the effectiveness of Bagging in enhancing predictive accuracy for various geospatial phenomena.

This model was chosen for this study due to its ability to reduce variance and enhance model stability, particularly in geographic modeling tasks, where data variability can significantly impact predictive performance. While Random Forest is another popular ensemble method frequently applied in geospatial analysis [24], [25], it is a variant of Bagging that focuses on constructing multiple decision trees. In contrast, Bagging can be applied with a variety of base learners, offering greater flexibility. Random Forest incorporates random feature selection at each split, which can be advantageous for certain types of data but may lead to higher bias when modeling complex spatial patterns.

Support Vector Machines (SVM) are also widely used in geographic information system (GIS) modeling due to their strong performance in high-dimensional spaces [26]. However, SVM can be sensitive to noise in the data and may require extensive parameter tuning, especially in cases where the data distribution is complex and non-linear. Bagging, by aggregating the predictions of multiple base models, mitigates the risk of overfitting and noise sensitivity, providing a more balanced approach to spatial data modeling.

2) Frequency Ratio (FR): The goal of analyzing the mountain hiking conditioning factors is to carefully assess how different geographic parameters relate to the presence of mountain hiking sites. To do this, we use the frequency ratio method, which allows us to measure how often hiking sites appear in different classes of geographic factors (such as elevation, slope, or land use) compared to how much area those classes cover. As explained by Manaouch et al. [23], this method helps determine the likelihood of finding a hiking site in specific geographic conditions.

The frequency ratio technique works by comparing two percentages: the percentage of mountain hiking sites located in a certain class and the percentage of the total area that this class occupies. A higher frequency ratio means there is a greater chance of finding a hiking site in that specific geographic class, as noted by Samanta et al. [27].

The frequency ratio for a particular class is calculated using the following equation:

$$FR_i = \frac{\left(\frac{N_i}{N}\right)}{\left(\frac{S_i}{S}\right)} \tag{3}$$

Where:

- N_i is the number of mountain hiking sites within the ith class of a geographic factor (e.g., a specific elevation range or slope category).
- N is the total number of mountain hiking sites in the study area.
- S_i is the area of the ith class of the geographic factor.
- S is the total area of the study region.

This formula helps us understand whether certain geographic conditions are more favorable for mountain hiking. For example, if a specific elevation range covers 20% of the study area but contains 50% of the hiking sites, the frequency ratio will be higher, indicating that this elevation is particularly suitable for hiking.

The frequency ratio equation is used to compute FR_i for each class of a conditioning factor, helping us evaluate how well different geographic conditions align with the locations of hiking sites.

C. Performance and Model Validation

The model's performance was evaluated using the Area Under the Curve (AUC) of the receiver operating characteristic (ROC) analysis. This analysis provides a comprehensive measure of the model's ability to distinguish between suitable and unsuitable mountain hiking areas. The AUC value ranges from 0 to 1, where a value of 0.5 indicates no discriminative power and a value of 1 signifies perfect discrimination.

For this study, the AUC/ROC analysis involved 120 mountain hiking sites and the classified map of potential hiking areas. True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) rates were calculated, which serve as critical metrics in evaluating the model's predictive accuracy. These metrics were derived using the following equations:

$$TPR = \frac{TP}{TP + FN} \tag{4}$$

$$FPR = \frac{FP}{FP + TN} \tag{5}$$

In this context, the True Positive Rate (TPR), also known as sensitivity or recall, reflects the proportion of actual positives that were correctly identified by the model. Conversely, the False Positive Rate (FPR) quantifies the proportion of actual negatives that were incorrectly classified as positives. By

analyzing these rates, we can better understand the trade-offs between sensitivity and specificity, which is crucial for applications where the consequences of false positives and false negatives differ significantly.

TABLE I. DATA LABELING FOR ROC ANALYSIS

Actual Label	Predicted Positive	Predicted Negative
Positive	TP	FN
Negative	FP	TN

Table I summarizes the data labeling used for the ROC analysis, indicating how actual labels correspond to predicted outcomes. This structured approach enables a clearer evaluation of the model's performance across different thresholds.

The resulting AUC value of approximately 0.93 indicates that the Bagging classifier performed exceptionally well in accurately identifying and distinguishing potential mountain hiking sites. Such validation not only affirms the reliability of the model but also enhances confidence among stakeholders in utilizing these findings for informed decision-making regarding tourism development.

Furthermore, the ROC curve, which graphically represents the TPR against the FPR at various threshold settings, provides an intuitive visualization of the model's performance. Analyzing the shape and area under this curve can reveal insights into the model's robustness, assisting in the selection of the optimal threshold that balances sensitivity and specificity according to the specific context of mountain hiking site selection.

D. Mountain Hiking Sites and Conditioning Factors

Table II presents the calculated frequency ratio values for the three conditioning factors, each consisting of multiple classes. To facilitate analysis, these values have been normalized using the normalize filter method, which transforms them into a range of 0.1 to 0.9. A normalized value closer to 1 indicates a stronger association between mountain hiking sites and the corresponding factor, while a value closer to 0 indicates a weaker association. The spatial relationship between the conditioning factors and mountain hiking sites is assessed through the frequency ratio calculation, and the results are presented in Table II for all factors.

E. Mountain Hiking Suitability Map

Fig. 8 illustrates the mountain hiking suitability map generated by the Bagging classifier. The map highlights that the northern and western regions exhibit higher suitability for potential mountain hiking sites compared to other areas. To gain a deeper understanding of the factors influencing the suitability or unsuitability of different areas and to explore the underlying reasons for these variations, we collaborated with local experts who possess substantial knowledge of the region.

Additionally, we validated our findings by cross-referencing them with images obtained from Google Earth. Fig. 10 provides the corresponding percentages of the area covered by each suitability class in the generated map.

TABLE II. SPATIAL RELATIONSHIP BETWEEN SLOPE AND MOUNTAIN CLIMBING SITES POINTS USING FREQUENCY RATIO (FR)

RCF	Class	Frequency Ratio FRi	Frequency Ratio FRn
Slope	1	0.045	0.1
	2	0.3	0.12
	3	0.74	0.15
	4	1.43	0.21
	5	9.58	0.9
Elevation	1	0.1	0.1
	2	0.2	0.1
	3	0.69	0.13
	4	0	0.1
	5	0.2	0.1
	6	0	0.1
	7	7.6	0.55
	8	4.17	0.34
	9	0	0.1
Length of slope	1	0	0.1
	2	0	0.1
	3	0.16	0.11
	4	0.5	0.24
	5	2.02	0.81

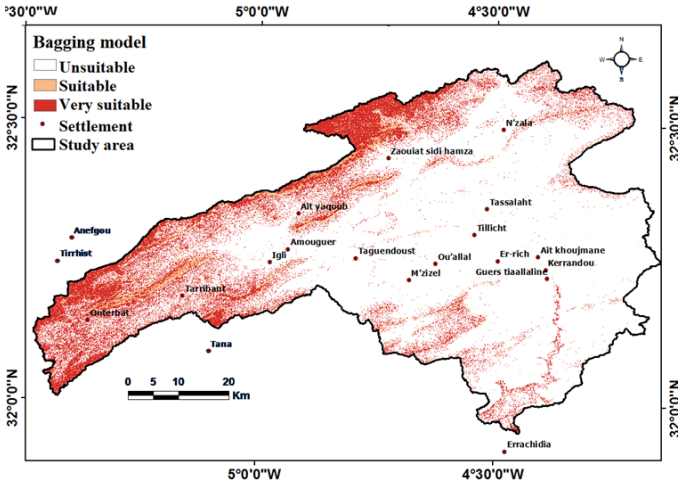


Fig. 8. Generated mountain hiking suitability map using Bagging.

F. Model's Performance

The effectiveness of the Bagging classifier was evaluated using the AUC/ROC method, which is widely recognized for its robustness in assessing model performance in binary classification tasks. During the training phase, the classifier demonstrated impressive performance, achieving an average accuracy rate of 97.29%. This high accuracy reflects the model's capability to learn from the training data effectively and underscores its potential in identifying suitable mountain hiking sites.

However, during the validation phase, significant fluctuations in the classifier's performance were observed, as depicted in Fig. 9. These fluctuations can be attributed to several factors, including variability in the validation dataset, potential overfitting during the training phase, and the inherent complexity of the mountain hiking suitability classification task. Understanding these fluctuations is crucial, as they highlight the necessity for careful model tuning and validation to ensure reliable outcomes in real-world applications.

The obtained AUC value of 0.935 indicates that the Bagging classifier successfully modeled the suitability mapping of mountain hiking sites in the upper Ziz region. An AUC value above 0.9 is generally considered excellent, suggesting that the model has a high probability of distinguishing between suitable and unsuitable hiking areas. This level of performance is particularly encouraging, as it implies that the classifier can reliably inform decisions regarding the identification of potential hiking locations.

Fig. 9 illustrates the prediction rate curve for the mountain hiking suitability map produced by the Bagging classifier. The curve visually represents the relationship between the true positive rate (TPR) and the false positive rate (FPR) across various threshold settings. Analyzing this curve allows for a deeper understanding of the trade-offs involved in selecting an optimal threshold, ensuring that the final suitability map meets the specific objectives of identifying suitable hiking sites.

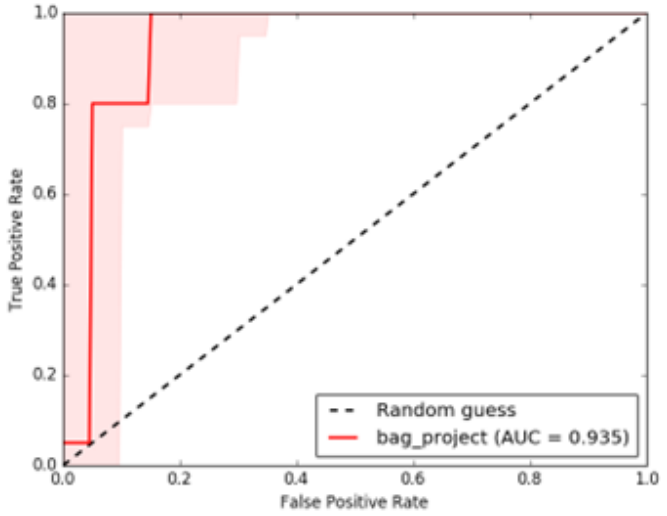


Fig. 9. Prediction rate curve for mountain hiking suitability map produced by Bagging.

IV. Discussion

A. Analysis of Precision

The performance of machine learning models (MLMs) can vary depending on the specific algorithm employed. In the case of Bagging, its performance was assessed using the 10-fold cross-validation method during the training phase, resulting in a high level of performance with an AUC value of 0.97.

However, during the validation phase, the performance of Bagging decreased to 93%, as depicted in Fig. 9. It is generally observed that algorithms with higher AUC values tend to exhibit more accurate and efficient prediction capabilities, as emphasized by Su et al. [28]. Considering the current AUC value of Bagging, it can be inferred that Bagging performs relatively lower compared to the results reported by Manaouch et al. [23] in predicting potential reforestation areas. Nevertheless, it is important to note that Bagging still demonstrates a favorable level of accuracy and efficiency, as evidenced by its AUC value of 0.93.

In evaluating the model's accuracy, it is essential to consider other factors that influence the suitability of regions for mountain hiking beyond the quantitative metrics provided by the model. For instance, cultural aspects, such as local traditions and community engagement in outdoor activities, can significantly impact the attractiveness of certain hiking locations. Furthermore, the presence of infrastructure, including access roads, trail maintenance, and proximity to amenities (like parking and rest areas), plays a crucial role in determining the feasibility and safety of hiking sites. Areas that may have high ecological or aesthetic value but lack necessary infrastructure may be less suitable for hiking compared to less pristine areas that are well-supported by facilities.

B. Distribution of Suitable Mountainous Hiking Areas

The data attributes collected from the upper Ziz region underwent preprocessing using the pre-trained Bagging classifier. During this preprocessing stage, mountain hiking suitability indices were computed for each pixel within the study area. The "Natural breaks" algorithm was employed to classify these suitability indices, resulting in the creation of three distinct classes: unsuitable, suitable, and very suitable. This process led to the generation of a mountain hiking suitability map, as depicted in Fig. 8.

The findings indicate that approximately 24% of the entire study area comprises very suitable areas for potential mountain hiking sites, as shown in Fig. 10. According to the Bagging model, the southern parts of the region are deemed unsuitable for mountain hiking sites, while the mountainous regions situated to the west and north exhibit a high level of suitability. Additionally, the Bagging model reveals that areas classified as very suitable for mountain hiking sites are dispersed along the Ziz wadi, located to the south of the upper Ziz region.

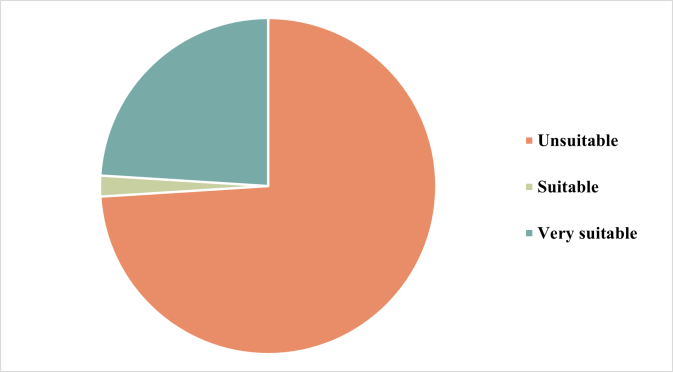


Fig. 10. Percentage of mountain hiking suitability classes' areas for Bagging model.

Despite these promising results, there are several limitations to this study that must be acknowledged. One significant limitation is the potential quality of the data used in the analysis. Incomplete or inaccurate data could adversely affect the model's predictive capabilities. Furthermore, the scope of topographical features included in the analysis may not encompass all relevant variables that could influence hiking suitability, such as microclimate variations, vegetation types, or the impact of human activities on the landscape. A broader

dataset could enhance the model's accuracy and provide a more comprehensive understanding of the factors affecting hiking suitability.

V. Conclusion and Future Work

The integration of topographical data and the Bagging algorithm has demonstrated significant potential in enhancing tourism through the identification of suitable mountain hiking sites. By utilizing the Bagging classifier, which incorporates critical topographical features such as slope, length of slope, and elevation, and validating the resulting map using the AUC/ROC metric, valuable insights can be gained to identify potential mountain hiking sites in the eastern High Atlas of the upper Ziz in Southeast Morocco. The findings indicate that approximately 24% of the study area comprises very suitable areas for potential mountain hiking sites, primarily located in the northern and western regions, as well as around the Ziz wadi, particularly in the downstream part of the study area.

The application of the AUC/ROC evaluation metric in this study provides a robust measure of the algorithm's ability to differentiate between suitable and unsuitable mountain hiking sites, achieving an impressive score of approximately 93%. This indicates that the Bagging algorithm performed exceptionally well in accurately identifying and distinguishing potential mountain hiking sites. Such validation approaches not only ensure the reliability and accuracy of the generated map but also empower decision-makers and stakeholders in the tourism industry to confidently identify and prioritize areas for tourism development.

Furthermore, the findings of this study underscore the importance of employing machine learning techniques in environmental and tourism planning. By systematically analyzing the spatial relationships among various topographical factors, stakeholders can make informed decisions that promote sustainable tourism practices while preserving natural landscapes.

Future research could explore the integration of additional ecological and socio-economic data to refine the suitability assessments further. Incorporating factors such as ecological diversity and human accessibility, including proximity to transport links and local amenities, can significantly enhance the accuracy of suitability models. Additionally, collaborating with local communities is crucial for validating findings. Engaging with residents can provide insights into cultural significance, traditional knowledge, and existing infrastructure, which are essential for a comprehensive understanding of the area's tourism potential.

Moreover, applying similar methodologies in other geographical areas could contribute to a broader understanding of potential hiking sites, ultimately supporting the growth of ecotourism on a larger scale. Future studies should also consider the impacts of climate change on hiking site suitability, ensuring that long-term strategies for tourism development are both adaptable and sustainable.

Overall, this study not only contributes to the literature on tourism but also offers practical applications for enhancing the visitor experience in mountain environments, fostering both environmental conservation and local economic development.

Acknowledgment

This research project was supported by funding from the CNRST (National Center for Scientific and Technical Research) under the Moroccan project titled "Al-Khawarizmi program in AI and its Applications", (Alkhawarizmi/2020/32).

The authors would like to express their gratitude to Mohamed Manaouch for providing the necessary facilities and data for this study. Additionally, the authors would like to extend their thanks to Mr. N. Khan for his invaluable assistance during the validation phase

References

[1] Y. Wang, L. Zhou, H. Chen, L. Wang, and X. Wu, "Framing memorable hiking tourism experiences through embodiment: The case of mount huangshan, china," *Journal of Outdoor Recreation and Tourism*, vol. 45, p. 100710, 2024.

[2] Y. Galiakbarov, O. Mazbayev, L. Mutaliyeva, V. Filimonau, and H. Sezerel, "When the mountains call: Exploring mountaineering motivations through the lens of the calling theory," *Journal of Outdoor Recreation and Tourism*, vol. 45, p. 100743, 2024.

[3] H. Naseri, P. Nourbakhsh, M. Kohandel, and A. Khodayari, "Factors affecting sustainable development in mountaineering," *Environmental Education and Sustainable Development*, vol. 12, no. 2, pp. 165–183, 2024.

[4] Z. Liu, T. Yang, C. Yi, and K. Zhang, "Effects and functional mechanisms of serious leisure on environmentally responsible behavior of mountain hikers," *Journal of Outdoor Recreation and Tourism*, vol. 45, p. 100709, 2024.

[5] C. M. Rogerson and J. M. Rogerson, "Historical special interest tourism: The evolution of mountaineering in south africa," *Revistă Română de Geografie Politică*, vol. 26, no. 1, pp. 1–13, 2024.

[6] N. Kchikach, H. Ibouh, A. Benali, A. Charbaoui, and A. E. A. El Fels, "Geoheritage of a mountain tourist route (high atlas, morocco): Digital promotion and visitor typology," 2024.

[7] L. Naimi, E. M. Bouziane, and A. Jakimi, "A model-driven approach for developing smart tourism web applications," in *International Conference on Digital Technologies and Applications*. Springer, 2024, pp. 255–263.

[8] L. Naimi, A. Jakimi, L. El Bermi *et al.*, "A sketch of a mda approach to accelerate the development of user interfaces in tourism web applications," in *2024 Mediterranean Smart Cities Conference (MSCC)*. IEEE, 2024, pp. 1–5.

[9] M. Manaouch, S. Mohamed, and F. Imad, "Coupling fuzzy logic and analytical hierarchy process (fahp) with gis for landslide susceptibility mapping (lsm) in ziz upper watershed, se morocco," 2021.

[10] M. Manaouch, M. Sadiki, M. Aghad, Q. B. Pham, M. Batchi, and J. Al Karkouri, "Assessment of landslide susceptibility using machine learning classifiers in ziz upper watershed, se morocco," *Physical Geography*, vol. 45, no. 2, pp. 203–230, 2024.

[11] M. Manaouch, M. Sadiki, and I. Fenjiro, "Integrating gis-based fahp and watem/sedem for identifying potential rwh areas in semi-arid areas," *Geocarto International*, vol. 37, no. 25, pp. 8882–8905, 2022.

[12] A. M. Al-Abadi, "Mapping flood susceptibility in an arid region of southern iraq using ensemble machine learning classifiers: a comparative study," *Arab. J. Geosci*, vol. 11, p. 218, 2018.

[13] P. Beedie and S. Hudson, "Emergence of mountain-based adventure tourism," *Annals of Tourism Research*, vol. 30, no. 3, pp. 625–643, 2003.

[14] R. Buckley, "Impacts of ecotourism on birds." 2004.

[15] G. Herrera-Franco, P. Carrión-Mero, N. Montalván-Burbano, J. Caicedo-Potosí, and E. Berrezueta, "Geoheritage and geosites: A bibliometric analysis and literature review," *Geosciences*, vol. 12, no. 4, p. 169, 2022.

[16] D. B. Weaver, "Ecotourism as mass tourism: Contradiction or reality?" *Cornell Hotel and Restaurant Administration Quarterly*, vol. 42, no. 2, pp. 104–112, 2001.

[17] K. Thapa, D. King, Z. Banhalmi-Zakar, and A. Diedrich, "Nature-based tourism in protected areas: a systematic review of socio-economic benefits and costs to local people," *International Journal of Sustainable Development & World Ecology*, vol. 29, no. 7, pp. 625–640, 2022.

[18] R. Huggett and E. Shuttleworth, *Fundamentals of Geomorphology*. Routledge, 2022.

[19] D. Huber, J. Freidl, C. Pichler, M. Bischof, M. Kiem, R. Weisböck-Erdheim, G. Squarra, V. De Nigris, S. Resnyak, M. Neberich *et al.*, "Long-term effects of mountain hiking vs. forest therapy on physical and mental health of couples: a randomized controlled trial," *International journal of environmental research and public health*, vol. 20, no. 2, p. 1469, 2023.

[20] A. Bizuwerk, G. Taddese, and Y. Getahun, "Application of gis for modelling soil loss rate in awash river basin," *Ethiopia, ILRI*, 2003.

[21] L. Breiman, "Using iterated bagging to debias regressions," *Machine Learning*, vol. 45, pp. 261–277, 2001.

[22] Q. Zhang, Z. Ning, X. Ding, J. Wu, Z. Wang, P. Tsangaratos, I. Ilia, Y. Wang, and W. Chen, "Hybrid integration of bagging and decision tree algorithms for landslide susceptibility mapping," *Water*, vol. 16, no. 5, p. 657, 2024.

[23] M. Manaouch, M. Sadiki, Q. B. Pham, A. Zouagui, M. Batchi, and J. Al Karkouri, "Predicting potential reforestation areas by quercus ilex (l.) species using machine learning algorithms: case of upper ziz, southeastern morocco," *Environmental Monitoring and Assessment*, vol. 195, no. 9, p. 1094, 2023.

[24] M. Geerts, S. vanden Broucke, and J. De Weerdt, "Georf: a geospatial random forest," *Data Mining and Knowledge Discovery*, pp. 1–35, 2024.

[25] H. Talebi, L. J. Peeters, A. Otto, and R. Tolosana-Delgado, "A truly spatial random forests algorithm for geoscience data analysis and modelling," *Mathematical Geosciences*, vol. 54, no. 1, pp. 1–22, 2022.

[26] K. Sun, Z. Li, S. Wang, and R. Hu, "A support vector machine model of landslide susceptibility mapping based on hyperparameter optimization using the bayesian algorithm: a case study of the highways in the southern qinghai–tibet plateau," *Natural Hazards*, pp. 1–22, 2024.

[27] S. Samanta, D. K. Pal, and B. Palsamanta, "Flood susceptibility analysis through remote sensing, gis and frequency ratio model," *Applied Water Science*, vol. 8, no. 2, p. 66, 2018.

[28] Q. Su, W. Tao, S. Mei, X. Zhang, J. Guo, and Y. Yang, "Landslide susceptibility zoning using c5.0 decision tree, random forest, support vector machine and comparison of their performance in a coal mine area," *Front. Earth Sci*, vol. 1181, 2021.

Pioneering Granularity: Advancing Native Language Identification in Ultra-Short EAP Texts

Zhendong Du, Kenji Hashimoto
Graduate School of Information, Production and Systems
Waseda University, Fukuoka, Japan 808-0135

***Abstract*—This study addresses the challenge of Native Language Identification (NLI) in ultra-short English for Academic Purposes (EAP) texts by proposing an innovative two-stage recognition method. Conventional views suggest that ultra-short texts lack sufficient linguistic features for effective NLI. However, we have found that even in such brief texts, subtle linguistic cues—such as syntactic structures, lexical choices, and grammatical errors—can still reveal the author's native language background. Our approach involves fine-tuning the granularity of first language (L1) labels and refining deep learning models to more accurately capture the subtle differences in second language (L2) English texts written by individuals from similar cultural backgrounds. To validate the effectiveness of this method, we designed and conducted a series of scientific experiments using advanced Natural Language Processing (NLP) techniques. The results demonstrate that models adjusted for granular L1 distinctions exhibit greater sensitivity and accuracy in identifying language variations caused by nuanced cultural differences. Furthermore, this method is not only applicable to ultra-short texts but can also be extended to texts of varying lengths, offering new perspectives and tools for handling diverse language inputs. By integrating in-depth linguistic analysis with advanced computational techniques, our research opens up new possibilities for enhancing the performance and adaptability of NLI models in complex linguistic environments. It also provides fresh insights for future efforts aimed at optimizing the capture of linguistic features.**

***Keywords*—Native language identification; English for academic purposes; natural language processing**

I. Introduction

NLI is a technique used to determine an author's native language by evaluating their written text. This method plays a crucial role in L2 writing research [1]. Traditionally, it has been assumed that effective NLI requires longer texts [2], as extended content provides a broader range of stylistic and linguistic features, making it easier to infer the author's native language with greater accuracy. This belief stems from the linguistic diversity and variability seen across different languages, as individuals' written expressions are shaped by their unique cultural and educational backgrounds [3]. However, these influences are often difficult to capture in shorter texts, leading to the widespread misconception that extremely short texts are unsuitable for NLI.

This assumption, however, overlooks certain fundamental aspects of linguistics. Even in very short texts, specific linguistic features—such as grammatical structures, lexical choices, and common errors—can reveal clues about the author's native language. For instance, writers from different language backgrounds often follow identifiable patterns when using prepositions, articles, or complex sentence structures [4]. These patterns may still be present even in short text segments. Moreover, mistakes such as misusing certain tenses or irregular verb forms can provide valuable insights for NLI.

In many practical applications, it is common to encounter texts at the sentence level or even shorter [5]. This reality has driven the exploration of NLI techniques capable of handling ultra-short texts, particularly in the context of EAP. In EAP writing, each word choice and grammatical structure may reflect the author's linguistic habits and native language influences [6]. Thus, even brief texts, such as titles, abstracts, or notes, can contain enough information to facilitate effective NLI.

Recognizing that traditional NLI approaches often overlook the subtle linguistic cues in ultra-short texts, this study proposes an innovative two-stage NLI method. By fine-tuning the granularity of NLI classification labels, we enhance the system's ability to analyze very short texts. This approach challenges conventional wisdom and proves particularly useful for academic English, enabling accurate identification of subtle differences in expression that arise from cultural and linguistic background variations.

Moreover, our method emphasizes the importance of linguistic analysis to uncover and utilize nuanced differences, such as slight variations in grammatical structures or lexical preferences. These aspects are often neglected in traditional NLI methodologies. To validate the effectiveness of our approach, we conducted a series of targeted experiments using advanced machine learning models. The results demonstrate that our method not only excels in identifying native languages from ultra-short texts but also offers more precise language analysis for applications like academic writing assistants and Grammatical Error Correction (GEC) tools. These applications can provide more tailored writing suggestions and GEC by accurately identifying the author's native language, ultimately improving both the quality and efficiency of writing. This study's success paves the way for further integration of linguistic research and natural language processing technologies.

II. Related Work

A. Overview

NLI aims to infer an author's native language based on the text written in a target language. This task holds significant importance in the field of NLP due to its various applications. For instance, NLI can be leveraged to improve language teaching methods, enhance the quality of machine translation, and bolster security monitoring capabilities. By understanding an

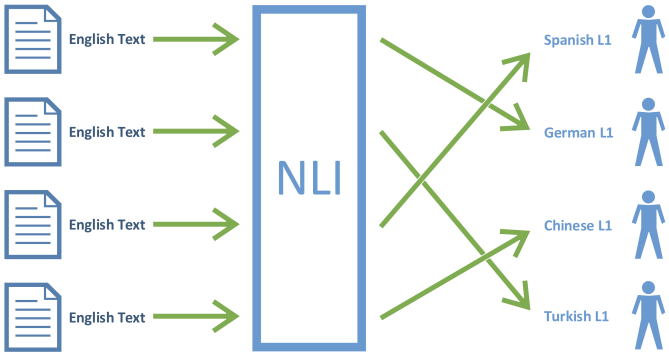


Fig. 1. The general concept of an NLI system is depicted in this figure. Image adapted from [7].

author's native language, we can more accurately identify their linguistic habits and potential errors [8], thereby providing valuable support for downstream tasks [9]. Fig. 1 illustrates the concept of NLI.

B. English NLI

English NLI is the most extensively studied area within NLI. This prominence is due to the widespread use and influence of English as the primary international language. In English NLI, researchers focus on extracting linguistic features from English texts that reflect the author's native language characteristics. Most existing studies have concentrated on long texts, such as essays [10], articles [11], and speeches [12]. These long texts provide a wealth of data, enabling models to capture the linguistic habits and preferences of speakers from different native language backgrounds. Common analytical dimensions include lexical usage frequency [13], syntactic structures [14], and pragmatic features [15].

Academic English NLI is a specialized subfield of NLI, aiming to infer the author's native language background through the analysis of academic English texts. Due to the relatively uniform style and conventions of academic English, extracting native language features poses a greater challenge [16]. Nevertheless, authors from different native language backgrounds exhibit variations in lexical choices [17], syntactic complexity [18], and argumentation styles within academic writing. Some studies utilize corpora comprising academic papers, research reports, and students' academic writings to analyze features such as discipline-specific terminology [20], the frequency of passive voice usage [21], and the overall text organization structure [22].

C. Short Text NLI

It is important to note that current research primarily focuses on NLI for long texts [23], while there is skepticism regarding the feasibility of effective NLI for very short texts, such as at the sentence level. This skepticism stems from the limited linguistic features available in ultra-short texts, making it challenging for models to capture stable native language traits [24]. Despite this, NLI for short texts is essential for specific NLP downstream tasks, such as English GEC or writing assistance tools. Unfortunately, due to the prevailing view that NLI has a minimum text length requirement, researchers have not yet explored this area adequately.

We argue that the conclusion regarding the limitations of NLI for short texts arises because prior studies did not consider different approaches for varying text lengths. Long texts, such as paragraphs, contain a wealth of linguistic features, and methods developed for them are often inappropriately applied to sentence-level texts. Hence, the current state of research in this area requires further investigation and reevaluation.

III. Goal

This study aims to address the gap in NLI research for very short texts by proposing a systematic and scientific solution. To ensure the quantifiability of the research, the focus is specifically on EAP, with the text length constrained to the sentence level.

IV. Method

NLI is a typical text classification task where the core challenge lies in enabling the NLI system to effectively learn to balance different classification labels [25]. Traditional classification methods often struggle when the NLI system is presented with sentence-level texts that have minimal distinguishing features. However, even at the sentence level, there are usually subtle differences present [26]; the key challenge is how to identify and leverage these differences.

A recent study explored this by extracting sentences from academic papers and translating them into different languages using a specific method [27]. These sentences were then translated back into English by English L2 speakers whose native languages matched the target language. The researchers found that texts produced by L2 speakers from similar cultural backgrounds exhibited certain similarities, which were evident in aspects such as grammatical errors and linguistic style. Fig. 2 illustrates part of the study's findings, showing the distribution of grammatical errors. Some researchers may view these results as reinforcing the idea that such linguistic similarities make sentence-level NLI even less feasible. However, we take an opposing stance.

We propose that, by dynamically adjusting the granularity of classification labels during the training of an NLI system, and allowing the system to learn the features of English texts in stages, it becomes possible to capture those subtle features more effectively. This approach may enhance the system's ability to identify nuanced linguistic characteristics that are otherwise overlooked in conventional classification methods.

Fig. 2 reveals some intriguing patterns in the distribution of grammatical error types among certain L2 learners. Notably, learners from China, Japan, and South Korea exhibit highly similar distributions of grammatical errors. According to the original authors, this similarity can be attributed to the cultural and linguistic ties shared by these three countries, which result in analogous challenges in learning English as a second language.

Our proposed approach involves training the model to first learn the shared linguistic features present in English texts produced by L2 learners from these three countries. Following this, the model will focus on the subtle, unique characteristics

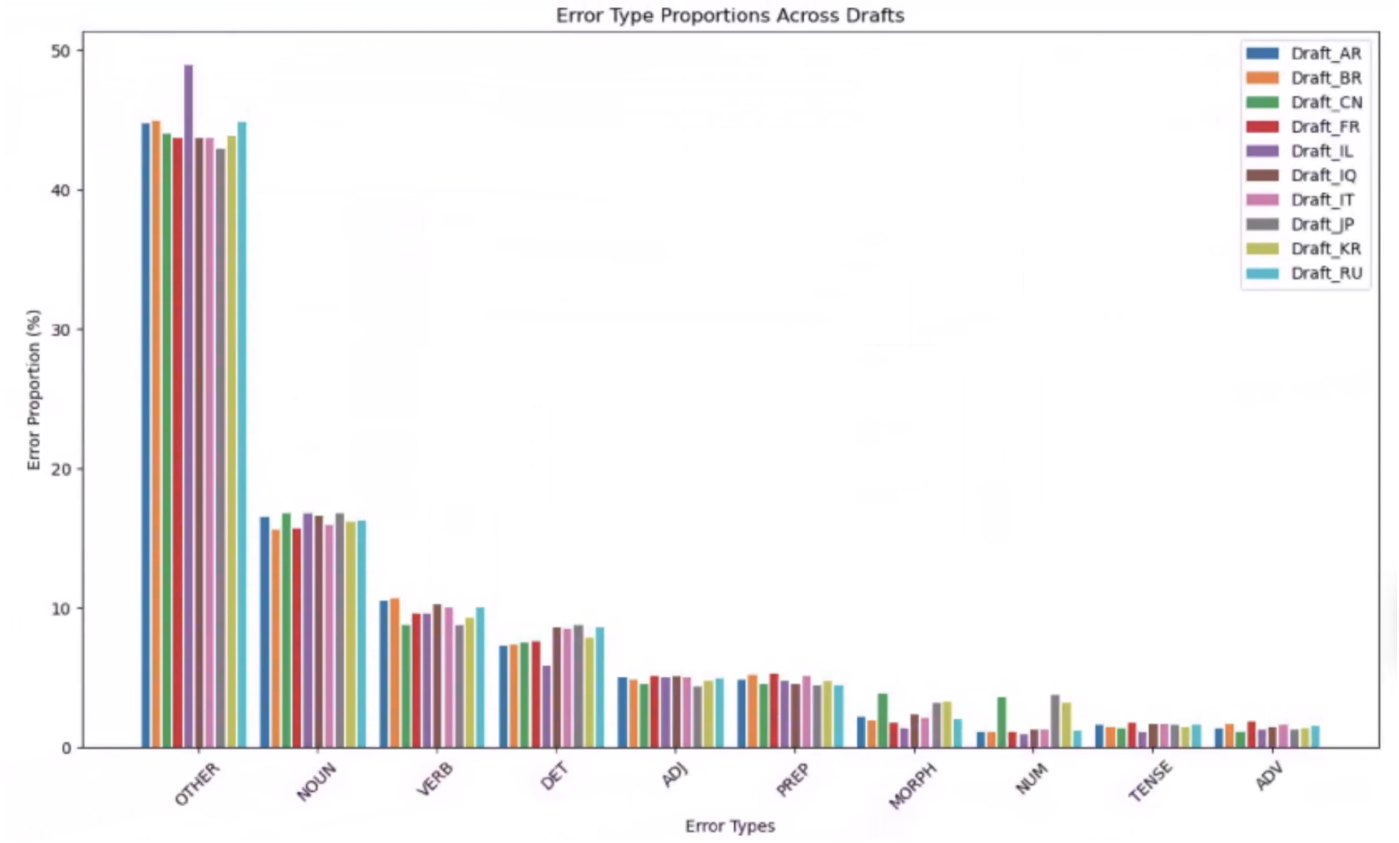


Fig. 2. Distribution of grammatical error types across different L2 transcribed texts. Different colors represent the native languages of various L2 speakers. Image adapted from [27].

of English usage specific to each country. Although similar patterns of error distribution may exist among L2 learners from other countries, our study is constrained by the availability of existing research data. Therefore, we have chosen to focus on learners from China, Japan, and South Korea. The specific methodology is detailed in Fig. 3.

V. Experiment

A. Data

In our study, we have selected the validation set of the TCNAEC [27] dataset as the training data and the test set as the testing data. To the best of our knowledge, there are currently no other publicly available datasets that meet the specific requirements of our task. While there are some English datasets consisting of paragraph collections that can be segmented into sentences, they lack the necessary labeled test data. Specifically, our task requires both the training and testing datasets to include English texts produced by L2 speakers from China, Japan, and South Korea. TCNAEC is the only dataset that satisfies this condition. It comprises 10 categories labeled as AR (Argentina), BR (Brazil), CN (China), FR (France), IL (Israel), IQ (Iraq), IT (Italy), JP (Japan), KR (South Korea), and RU (Russia), with each label containing 1,000 validation entries and 1,000 test entries.

B. Model

In this study, we selected the RoBERTa-large [28] model as the pre-trained foundation for our experiments. RoBERTa-large was chosen due to its robust text classification capabilities, which makes it well-suited to validate the proposed approach. The experiment was divided into two phases, each with distinct objectives and configurations.

In the first experiment, we followed the conventional method where data, balanced across original labels, was input into the model. This model is referred to as RoBERTa-10[1]. For the second experiment, we implemented our proposed method, where the three labels—CN, JA, and KR—were initially merged into a single label, CJK. This produced an intermediary model, referred to as NLI Model 1. Subsequently, using the same settings as in the first experiment, the data was reclassified into CN, JA, and KR labels, resulting in NLI Model 2, named RoBERTa-8to10[2].

The two experiments employed distinct hyperparameter configurations across different phases of training. In the first experiment, parameters were set according to configuration two (shown in Table II). For the second experiment, the first phase used configuration one (Table I), while the second phase used configuration two. The rationale behind using different parameter configurations for each phase was to address the

[1]For more details about the **RoBERTa-10**, visit **this link**.
[2]For more details about the **RoBERTa-8to10**, visit **this link**.

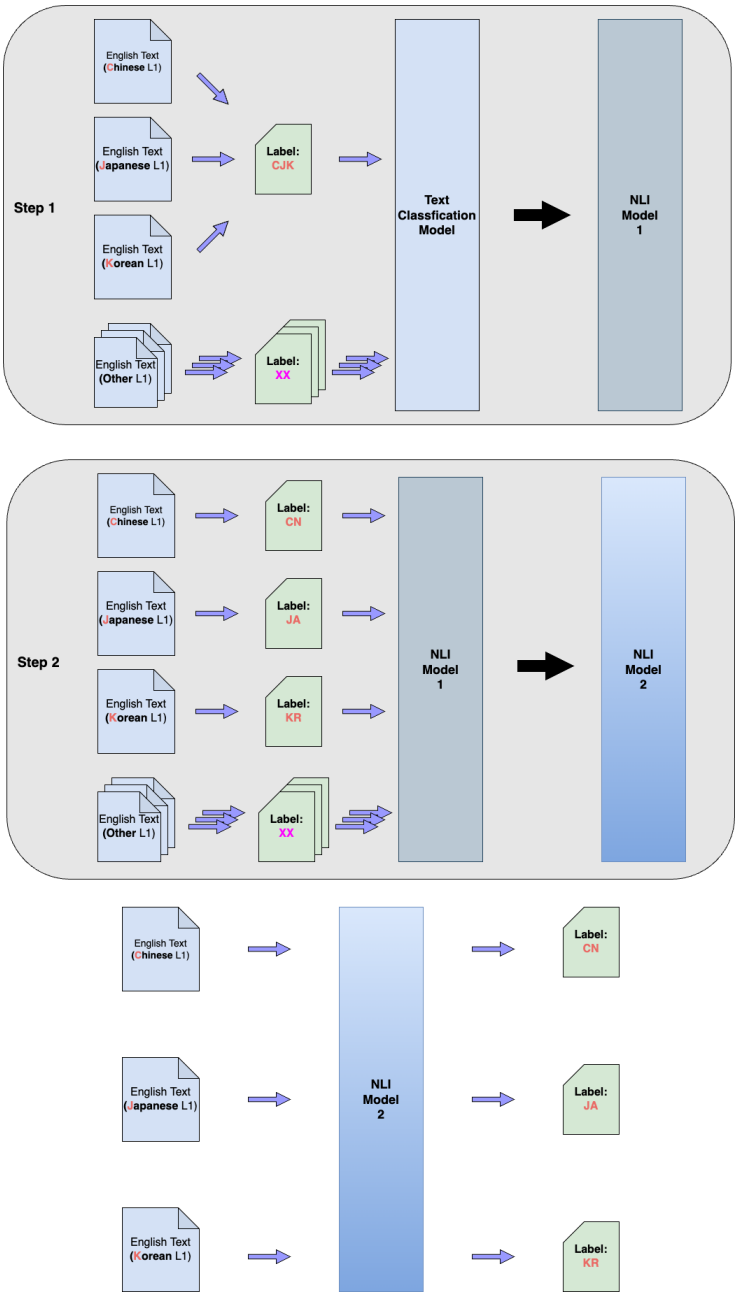


Fig. 3. Method details.

Step 1 Consolidates English text from Chinese, Japanese, and Korean L1 into the ***CJK*** label and feeds it, along with other labels, into a Text Classification Model to create **NLI Model 1**.

Step 2 Re-divides the ***CJK*** label into ***CN***, ***JA***, ***KR*** labels, which, alongside other labels, are fed into **NLI Model 1** to generate **NLI Model 2**. **NLI Model 2** can then more accurately identify and categorize English text from Chinese, Japanese, and Korean L1.

varying demands of the tasks and differences in the distribution of data labels.

Step 1: Establishing Baseline Classification Capabilities: The objective of the first phase was to establish NLI Model 1, which was designed to handle tasks involving broader categorical distinctions. Specifically, the model's primary goal was to classify English text originating from Chinese, Japanese, and Korean into the unified CJK label, alongside other distinct labels. This strategy aimed to capture key distinguishing features across broader language groups while ignoring minor linguistic variations, thereby forming a foundational understanding and classification ability for the main language groups.

To accommodate this requirement, we opted for a relatively large batch size (32) and a high learning rate (0.0001) to facilitate rapid convergence in the early stages of training. A cosine learning rate scheduler with restarts was employed to help the model effectively navigate potential local minima during training. Additionally, we adjusted the class weights to reflect label imbalances, particularly assigning a lower weight (0.333) to the CJK class to prevent the model from dispropor-

tionately favoring this category, which combined texts from three different languages.

Step 2: Enhancing Language Recognition Precision: In the second phase, the focus shifted towards more fine-grained language recognition. The objective of RoBERTa-8to10 in this phase was to differentiate between CN, JA, and KR labels, building on the language recognition capabilities developed in the first phase. To capture the nuanced differences between these languages, we adopted a smaller batch size (16) and a lower learning rate (0.00005), encouraging the model to become more sensitive to subtle details and achieve higher accuracy. To further prevent overfitting and ensure better generalization across fine-grained categories, we increased the dropout rate to 0.5.

By employing this step, targeted training strategy with parameter adjustments tailored to each phase's needs, RoBERTa-8to10 was able to improve classification accuracy and sensitivity when distinguishing between Chinese, Japanese, and Korean texts. This approach successfully met the distinct objectives of each phase.

TABLE I. HYPERPARAMETER SETTING 1

Parameter	Value
Mixed Precision	none
Optimizer	adamw_torch
Scheduler	cosine_with_restarts
Batch Size	32
Epochs	20
Gradient Accumulation	1
Learning Rate	0.0001
Maximum Sequence Length	128
Dropout Rate	0.3
Class Weight	AR: 1 BR: 1 CJK: 0.333 FR: 1 IL: 1 IQ: 1 IT: 1 RU: 1

TABLE II. HYPERPARAMETER SETTING 2

Parameter	Value
Mixed Precision	none
Optimizer	adamw_torch
Scheduler	cosine_with_restarts
Batch Size	16
Epochs	30
Gradient Accumulation	1
Learning Rate (lr)	0.00005
Maximum Sequence Length	128
Dropout Rate	0.5
Class Weight	balanced

VI. RESULTS

A. Analysis

We compared the classification performance of two models, RoBERTa-10 and RoBERTa-8to10, through metric evaluations (Table III) and confusion matrix analyses (Fig. 4) to explore the impact of a phased classification strategy on model performance.

TABLE III. MODEL PERFORMANCE METRICS

Model	Metric	CN	JP	KR
RoBERTa-10	Precision	28.02	35.29	10.12
	Recall	**78.60**	1.20	2.50
	$F_{0.5}$	32.16	5.28	6.29
	Average $F_{0.5}$		14.58	
RoBERTa-8to10	Precision	**33.06**	**42.37**	**16.85**
	Recall	59.50	**2.50**	**16.80**
	$F_{0.5}$	**36.28**	**10.11**	**16.84**
	Average $F_{0.5}$		**21.08**	

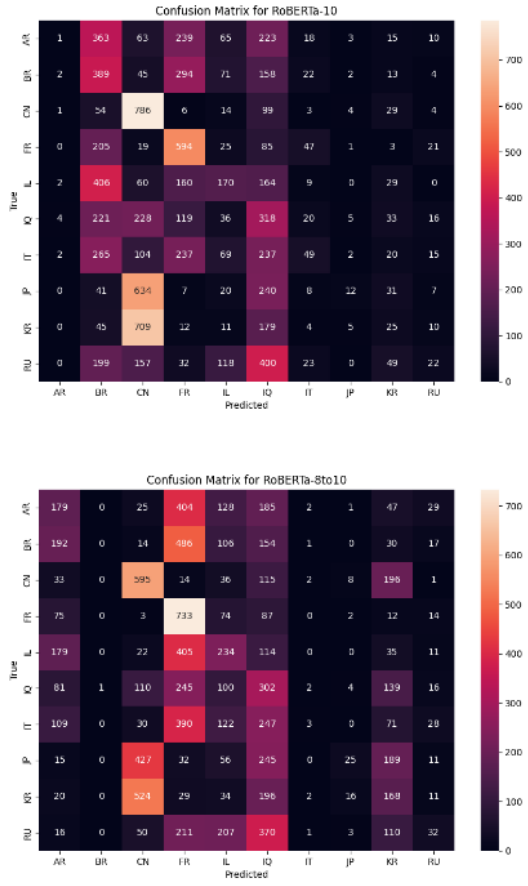


Fig. 4. Confusion matrix for models.

In the RoBERTa-10 model, all 10 labels were used simultaneously for training and prediction. The metric evaluations show that the model achieved a relatively high recall rate (78.60) for the CN label, but its precision was low (0.2802), resulting in a relatively low $F_{0.5}$ (32.16). This suggests that although the model successfully captured a significant portion of CN-labeled samples, its accuracy was insufficient, with a high number of misclassifications. The confusion matrix further confirmed this: while 786 CN samples were correctly classified, a considerable number of samples were misclassified as JP or KR. For the JP and KR labels, the recall rates were extremely low, with only a small number of samples correctly classified, and most were misclassified as CN. This indicates that the model struggled significantly in distinguishing between

these three labels.

In contrast, the RoBERTa-8to10 model employed a phased classification strategy. First, the CN, JP, and KR labels, which were prone to confusion, were merged into a new CJK label. This CJK label, along with the remaining seven labels, was used for initial model training. In the second phase, samples predicted as CJK were further subdivided. Metric evaluations showed an improvement in precision for the CN label to 33.06, with a corresponding increase in the $F_{0.5}$ score to 36.28, despite a slight drop in recall (0.5950). This indicates that the model effectively reduced misclassifications while maintaining a good ability to identify CN-labeled samples. For the JP and KR labels, both precision and $F_{0.5}$ scores saw significant improvements, particularly for the JP label, where precision increased from 35.29 to 42.37. The confusion matrix revealed that the number of correctly classified JP samples doubled to 25, while the number of correctly classified KR samples surged to 168. This highlights the effectiveness of the phased classification strategy in enhancing the model's ability to differentiate between these confusing labels [29].

In summary, the RoBERTa-8to10 model demonstrated substantial improvements in classification performance for the CN, JP, and KR labels. By merging the easily confused labels in the initial phase, the phased classification strategy reduced the complexity of the initial classification task. In the subsequent fine-tuning phase, the model was able to better learn the subtle differences between these labels, leading to enhanced precision and reliability. These results validate the advantages of a phased classification approach in multi-label classification tasks and provide an effective solution for addressing similar classification challenges.

B. Discussion and Future Work

In this work, we have demonstrated the effectiveness of the proposed method, showing that the phased classification strategy has significant advantages in handling multi-label classification tasks with easily confused labels. Due to limitations in data volume and the available types of data, our current exploration is restricted to English texts written by native speakers of Chinese, Japanese, and Korean. The linguistic and cultural similarities among these three languages make the classification task more challenging, but also provide an ideal testing ground for our model.

Looking ahead, we aim to address the issue of data scarcity. By collecting and constructing larger, more diverse datasets, we hope to validate our approach across a broader range of languages and cultural contexts. This would not only improve the generalization capabilities of the model but also allow us to explore the application of the phased classification strategy to more complex groupings and finer-grained labels.

Additionally, while this study focuses on sentence-level text classification, our approach is theoretically applicable to longer texts as well. For example, texts at the paragraph or full-article level contain richer contextual and semantic information [30], which could enable the model to more accurately capture linguistic features and distinctions. Thus, we plan to extend our method to longer texts in future research, with the expectation of achieving even better classification performance.

VII. Conclusion

In this study, we focus on NLI for ultra-short EAP texts and propose an innovative methodological strategy. This strategy primarily involves the phase-wise adjustment of L1 label granularity to fine-tune deep learning models, thereby more accurately capturing the subtle linguistic features of L2 English texts from similar cultural backgrounds. The core of this method integrates in-depth grammatical analysis with advanced computational techniques, combining traditional linguistic knowledge with modern machine learning technologies to enhance model performance and adaptability in complex linguistic environments. To validate our approach, we designed a series of scientific experiments that clearly demonstrate the effectiveness of the proposed method by comparing model performance before and after the experiments. The results show that the granularity-adjusted models are more sensitive to and can better identify linguistic variations caused by subtle cultural differences, thereby significantly improving the accuracy of NLI. Moreover, we conducted multiple rounds of verification to ensure the reliability and repeatability of the results. Through exhaustive data analysis, we not only proved the effectiveness of our method but also explored its potential for future research. The analysis indicates that this strategy is applicable not only to ultra-short texts but can also be effectively applied to texts of varying lengths, offering new perspectives and tools for managing diverse linguistic inputs. Our research further reveals that further optimization of the label granularity adjustment strategy could allow for more precise capture of linguistic features, providing new directions and possibilities for future research.

Acknowledgment

This work was supported by the Information, Production and Systems Research Center, Waseda University; the Future Robotics Organization, Waseda University; the Humanoid Robotics Institute, Waseda University, under the Humanoid Project; and the Waseda University Grant for Special Research Projects (grant number 2024C-518).

References

[1] S. Malmasi and M. Dras, "Multilingual native language identification," Natural Language Engineering, vol. 23, no. 2, pp. 163-215, 2017.

[2] B. Berti, A. Esuli, and F. Sebastiani, "Unravelling interlanguage facts via explainable machine learning," Digital Scholarship in the Humanities, vol. 38, no. 3, pp. 953-977, 2023.

[3] B. Berti, A. Esuli, and F. Sebastiani, "Unravelling interlanguage facts via explainable machine learning," Digital Scholarship in the Humanities, vol. 38, no. 3, pp. 953-977, 2023.

[4] J. Geluso, "Grammatical and functional characteristics of preposition-based phrase frames in English argumentative essays by L1 English and Spanish speakers," J. Engl. Acad. Purp., vol. 55, p. 101072, 2022.

[5] S. Steinbakken and B. Gambäck, "Native-language identification with attention," in Proc. 17th Int. Conf. Natural Language Processing (ICON), Dec. 2020, pp. 261-271.

[6] L.J. Zhang and X. Cheng, "Examining the effects of comprehensive written corrective feedback on L2 EAP students' linguistic performance: A mixed-methods study,"J. English for Academic Purposes, vol. 54, p. 101043, 2021.

[7] S. Malmasi et al., "Native language identification: explorations and applications," Ph.D. dissertation, Dept. of Sci. and Eng., Macquarie Univ., Sydney, Australia, 2016.

[8] J. A. Linck, J. F. Kroll, and G. Sunderman, "Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning," Psychological Science, vol. 20, no. 12, pp. 1507-1515, 2009.

[9] J. Tetreault, D. Blanchard, and A. Cahill, "A report on the first native language identification shared task," in Proc. 8th Workshop on Innovative Use of NLP for Building Educational Applications, Jun. 2013, pp. 48-57.

[10] S. Latifi, S. Syed, and M. Gierl, "Automated scoring of junior and senior high essays using Coh-Metrix features: Implications for large-scale language testing," Language Testing, vol. 38, no. 1, pp. 62-85, 2021.

[11] B. Gyawali, G. Ramirez, and T. Solorio, "Native language identification: a simple n-gram based approach," in Proc. Eighth Workshop on Innovative Use of NLP for Building Educational Applications, Atlanta, GA, USA, Jun. 2013, pp. 224-231.

[12] J. F. Werker and R. C. Tees, "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," Infant Behavior and Development, vol. 7, no. 1, pp. 49-63, 1984.

[13] S. Crossley, et al., "Frequency effects and second language lexical acquisition: Word types, word tokens, and word production," Int. J. Corpus Linguistics, vol. 19, no. 3, pp. 301-332, 2014.

[14] S. A. Kotz, "A critical review of ERP and fMRI evidence on L2 syntactic processing," Brain and Language, vol. 109, no. 2-3, pp. 68-74, 2009.

[15] G. Kasper, "Four perspectives on L2 pragmatic development," Applied Linguistics, vol. 22, no. 4, pp. 502-530, 2001.

[16] Y. K. Muthusamy, E. Barnard, and R. A. Cole, "Automatic language identification: A review/tutorial," IEEE Signal Processing Magazine, vol. 11, no. 4, pp. 33-41, Oct. 1994.

[17] S. A. Crossley and D. S. McNamara, "Computational assessment of lexical differences in L1 and L2 writing," Journal of Second Language Writing, vol. 18, no. 2, pp. 119-135, 2009.

[18] L. Ortega, "Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing," Applied Linguistics, vol. 24, no. 4, pp. 492-518, 2003.

[19] UYSAL, Hacer Hande. Argumentation across L1 and L2 Writing: Exploring Cultural Influences and Transfer Issues. Vigo International Journal of Applied Linguistics, 2012, 9: 133-159.

[20] P. Durrant, "Discipline and level specificity in university students' written vocabulary," Applied Linguistics, vol. 35, no. 3, pp. 328-356, 2014.

[21] Z. Du and K. Hashimoto, "Decoding Academic Language: The Symbiotic Relationship Between Boosters, Hedges, and Voice in EAP," in 2024 12th International Conference on Information and Education Technology (ICIET), 2024, pp. 46-52.

[22] Z. Du and K. Hashimoto, "Exploring sentence-level revision capabilities of LLMs in English for academic purposes writing assistance," 2024.

[23] M. Lui and P. Cook, "Classifying English documents by national dialect," in Proc. Australasian Lang. Technol. Assoc. Workshop 2013 (ALTA 2013), 2013, pp. 5-15.

[24] R. Zheng, J. Li, H. Chen, and Z. Huang, "A framework for authorship identification of online messages: Writing-style features and classification techniques," J. Am. Soc. Inf. Sci. Technol., vol. 57, no. 3, pp. 378-393, 2006.

[25] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," arXiv preprint arXiv:1801.06146, 2018.

[26] Z. Du and K. Hashimoto, "Data augmentation for sentrev using back-translation of lexical bundles," in Proc. 37th Pacific Asia Conf. Lang., Inf. Comput., Dec. 2023, pp. 70-79.

[27] Z. Du and K. Hashimoto, "Tcnaec: Advancing sentence-level revision evaluation through diverse non-native academic english insights," IEEE Access, vol. 2023.

[28] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," arXiv:1907.11692, 2019. [Online]. Available: https://arxiv.org/abs/1907.11692

[29] A. Tharwat, "Classification assessment methods," Applied Computing and Informatics, vol. 17, no. 1, pp. 168-192, 2021.

[30] K. Stock and J. Yousaf, "Context-aware automated interpretation of elaborate natural language descriptions of location through learning from empirical data," International Journal of Geographical Information Science, vol. 32, no. 6, pp. 1087-1116, 2018.

Feature Creation to Enhance Explainability and Predictability of ML Models Using XAI

Waseem Ahmed
Faculty of Computing and Information Technology
King Abdulaziz University
Jeddah, Saudi Arabia

***Abstract*—Bringing more transparency to the decision making process in fields deploying ML tools is important in various fields. ML tools need to be designed in such a way that they are more understandable and explainable to end users while bringing trust. The field of XAI, although a mature area of research, is increasingly being seen as a solution to address these missing aspects of ML systems. In this paper, we focus on transparency issues when using ML tools in the decision making process in general, and specifically while recruiting candidates to high-profile positions. In the field of software development, it is important to correctly identify and differentiate highly skilled developers from developers who are adept at only performing regular and mundane programming jobs. If AI is used in the decision process, HR recruiting agents need to justify to their managers why certain candidates were selected and why some were rejected. Online Judges (OJ) are increasingly being used for developer recruitment across various levels attracting thousands of candidates. Automating this decision-making process using ML tools can bring speed while mitigating bias in the selection process. However, the raw and huge dataset available on the OJs need to be well curated and enhanced to make the decision process accurate and explainable. To address this, we built and subsequently enhanced a ML regressor model and the underlying dataset using XAI tools. We evaluated the model to show how XAI can be actively and iteratively used during pre-deployment stage to improve the quality of the dataset and to improve the prediction accuracy of the regression model. We show how these iterative changes helped improve the r2-score of the GradientRegressor model used in our experiments from 0.3507 to 0.9834 (an improvement of 63.27%). We also show how the explainability of LIME and SHAP tools were increased using these steps. A unique contribution of this work is the application of XAI to a very niche area in recruitment, i.e. in the evaluation of performance of users on OJs in software developer recruitment.**

***Keywords*—XAI; ML; AI; Recruitment**

I. Introduction

The use of classification and regression models based on Machine Learning and Deep learning techniques in various domains is now common and ubiquitous. Advances in the field have brought faster diagnosis to patients undergoing life critical treatments in the medical domain, has made autonomous driverless vehicles possible and has revolutionized the commercial and financial world through prediction and recommendation systems. Although these systems are becoming increasingly pervasive in many domains, their adoption has become challenging in areas where trust, accountability and transparency of the decision making process is important and crucial to the users and other stakeholders of the application. One of the main reasons is that most of these ML models are closed *black-boxes* - they lack interpretability and explainability. ML models exist on both side of the spectrum. Models based on linear and logistic regression and classification are simple to understand and decisions made by them are easily explainable; but they lack accuracy and performance. On the other hand, there are ML models built on deep learning technologies which have great accuracy and performance but are not interpretable and decisions made by them are not explainable.

Besides this, there are other inherent problems in these systems. For example, they are trained on existing datasets and tested on an unseen dataset. However, this unseen data is, in reality, not really unseen but available to the developers at development and pre-deployment time. If there is a problem in the data or the data is biased in some way, the model is trained on this incorrect data and decisions it subsequently makes would be biased. And because these systems are trained on a specific data structure, they cannot be generalized or easily extended to systems where structure of data may be different or unknown.

In the early days of AI, systems were knowledge based [1] built on strong formalisms, rules and symbolic representations and thus decisions taken by them could, in most cases, be explained in detail. AI systems today, that are based on subsymbolic representations are skill based [1]. The focus of these systems in largely on performance and lesser on explainability. But for a ML system to be adopted in a life-critical or similar domains the system has to be considered trustworthy and transparent by the domain expert or the end user.

The field of Explainable AI (XAI) has been addressing the issue of bringing more transparency, understandability and explainability to these closed ML models [2], [3], [4], [5]. Bringing trust and explainability to ML systems while eliminating bias in decision making is critical in fields like Human Resources [6].

Systems based on ML are increasingly being used in the business world in decision making, which is sometimes leading to the decision making process being less transparent. Managers may not have the necessary skills to understand the decision making skills of the underlying black-box ML models used to make the decision (prediction or classification) and the XAI system should help them communicate with their team in a natural and more understandable way. XAI also seeks to addresses this concern to ensure that relationships between managers and other decision makers in such domains, which

are based on trust and transparency, are upheld. Ideally, if these systems are deployed in areas like HR, they should be able to bring accountability and address the big issue of eliminating bias during recruitment [7]. Providing a detailed justification when recruiting for a high-profile job may take a legal form and if aided with ML models, and a human understandable explanation will be sought as to why a candidate was accepted or rejected. Making algorithmic decision-making more accountable has been made a part of the Right to Explanation in the European Union General Data Protection Regulation (GDPR) [8].

Recently, HR recruitment in the software industry is increasingly being done using Online Judges through hackathons and online contests [9]. They are increasingly being used by companies to facilitate recruitment at different levels of skills ranging from software developers for various IT projects [10] to freshers in campus recruitment drives. Indeed, there are dedicated OJs primarily used to support the recruitment process like *CodeEval*, *Codility*, *HackerRank*, *HackerEarth* and *Qualified*. Data on these OJs is automatically generated with zero human intervention and is thus devoid of biases related to ethnicity, sex and color, thus, presenting a great opportunity to fields like recruitment where such discrimination is to be avoided.

In this research we show how XAI tools can be better used to bring more understandability and explainability to the recruitment process using OJs. Specifically, we show how XAI tools can be used for two main purposes during pre-deployment:

1) to understand and improve the performance of the ML model.
2) to improve the structure of the underlying dataset such that it contributes to increase the understandability and explainability of the ML model.

We designed and conducted iterative experiments to address both of the above aspects. Our experiences from these experiments are shared in the paper.

The rest of the paper is organized as follows. Section II provides a background on Online Judges (OJs) and XAI in the ML pipeline. Section III describes the first setup of the experiment. Section IV describes how domain knowledge is used to improvise both the model and the dataset. Observations and analysis using XAI on the extended and modified dataset are given in Section V. Related work is given in Section VI followed by the Conclusion section.

II. Related Work

Explainable Artificial Intelligence has drawn a lot of attention recently with lots of survey papers presented in literature on the topic [2], [4], [11], [3] and has drawn a lot of citations. It's importance and significance can be seen by DARPA's increasing interest in the field [12], [13]. Also, there are various domain specific papers that highlight the use and importance of XAI in critical fields like pathology, cancer detection, AI aided Alzheimer's Disease detection and bioinformatics [1], [14], [15], [5], [16]. It has also been used in non life-critical fields like Credit Risk Analysis [17] and in Human Resources in predicting employee attrition and implementing effective employee retention strategies [6]. Although the field of XAI has drawn a lot of attention recently creating new tracks within conferences and workshops and initiating new research initiatives and groups [2], it is not new. Researchers have been working in the field since many decades as detailed in [2].

AI and XAI have recently been used by HR in an increasing number of companies attempting to automate the recruitment process. ML has been used for screening resumes to shortlisting candidates, to analyzing employee churn, implementing strategies to improve employee retention, to identifying training and development needs, and in designing personalized development programs for employees. Some of these approaches have been evaluated by researchers and external agencies for their applicability. For example, [18] has applied the Design Thinking approach to evaluate the explainability aspect of a recruitment system by presenting counterfactuals generated by XAI systems [18]. A group of handpicked volunteers were selected for their experiment. However, their system is pretty basic and considers just two generic personas for evaluation and a relatively small data set.

XAI has also been adopted in fields closely related to software development. For example, in the software development process [19], in defect monitoring and prioritization [20], in software analytics [21] and during software maintenance [22]. There has been ongoing research in predicting developer performance with and without using AI and XAI principles. For example, the work done by [23] closely reflects our work in predicting developer performance without using XAI tools or techniques. Another similarity with our work is that it uses data from Codeforces for analysis. They train four different neural network models and obtain the best results for LSTM with Attention. This work differs from ours in many ways. Firstly, the dataset they use comprises of just 100 users. The paper does not explain how these 100 users were selected from about 140k registered users. The dataset seems too small to be representative of the different categories of users on Codeforces. Secondly, predictability of these users in terms of their behavior between contests is relatively consistent. Also, The paper does not consider temperamental performances of users in contests. Also, in their research, they consider past performances of users in both contests and practice sessions after the contest to predict rating after the contest while we use just contest data.

The work done in this paper differs from other previous works in two main aspects. First, we use XAI to both understand and improve the performance and explainability of the ML model. Second, we apply it to a system to predict developer performance in OJs. To the best of our knowledge, no work exists in this field. Moreover, the research work presented in this paper has a very narrow focus, i.e. evaluation of developer performance prediction during software development which may be in the form of a hackathon session created primarily for recruitment.

III. Background

A. Online Judges

Online Judge (OJ) platforms like SPOJ, EulerProject, Codeforces [24], Codechef, Topcoder, HackerEarth, etc. are being used by a wide spectrum of developers across the world

today to practice, sharpen and showcase their developmental skills. They are being used as development platforms, as crowd-sourcing platforms to hunt for solutions to industrial and science-driven challenges, as online compilers, in education, in workforce development and training, and more recently in recruitment [10] and by coaches in Universities to train and pick the best candidates to participate in the prestigious ICPC Challenge drawing more than 50,000 students from over 3000 Universities worldwide [25].

Most of them regularly hold online rated contests of fixed time durations in which thousands of users across the globe participate. Contestants are expected to solve a certain number of problems in each contest and are awarded *ranks*, *ratings* and *positions* after every contest and these are reflected on the OJ's leaderboard. These are described below in more detail.

1) Problems and verdicts: Each rated contest on an OJs presents contestants with problems at various difficulty levels ranging from elementary (easy) to hard (complex) [26]. In the context of this paper, we denote the difficulty level of a problem with θ ranging from 1 to 5 where θ=1 represents the difficulty level of the easiest problem in the contest and $\theta = 5$ represents the difficulty level of the hardest problem in the contest. Each problem is accompanied by a clear, and unambiguous description statement, sample inputs with their corresponding expected outputs, and time, memory and resource constraints under which the solution to the problem is to be executed to be successfully accepted by the OJ. Simple problems can be implemented with the use of simple logic and simple data structures while hard problems require the use of more sophisticated algorithms and efficient data structures in producing complete and correct solutions in addition to optimization to meet the specified performance and resource constraints.

For each submission made by a user, the OJ returns a verdict which can take one of the values specified in the first column of Table I.

2) Rating and divisions: When users participate in rated contests, they see a positive or negative change in rating that reflects their ability to solve the problems (tasks) in that contest relative to the previous contests. According to the rating, the contestants are split into multiple divisions based on their current rating. Division 1 on Codeforces consists of users who have the highest rankings (1900+), and problems in rated contests at this Division are generally the hardest to solve.

B. XAI Phases in a ML Pipeline

Consider a typical ML Pipeline indicated in Fig. 1 showing the two main phases where the use of XAI tools and techniques can be productive. The first phase is the *Understanding phase*, which subsumes the pre-deployment phases and comprises of Feature Engineering, Model Training and Model Testing. The second phase, is the *Explaining phase*, which subsumes the later two or more post-deployment phases.

The level of detail in the XAI generated explanations vary between the phases. For example, the explanation in the *Understanding* phase can be technical and given in semantics and formalism used by statisticians, mathematicians and informaticians. The first phase involves XAI tools being used

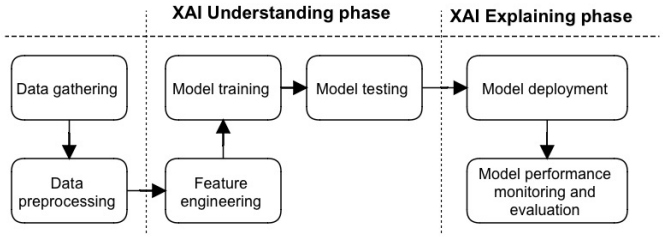


Fig. 1. A typical ML pipeline indicating XAI phases.

by the developer to improve the model during the training and testing stages and ensure that it works as intended [4] when it is deployed in the real-world. In the *Explaining phase*, however, the preferred way of explanation would be in Natural Language and/or intuitive and simple-to-understand visualizations. This phase would involve XAI tools being used by domain experts (a physician or a recruiting agent, for example) and end-users (a patient or a candidate for a job, for example) to interpret and better understand the reasons behind the decisions taken by the ML model deployed in real-life applications [4].

IV. ML Model with Traditional Stats

1) Data collection: For this research, we used *Codeforces* as the choice of OJ as they have easily accessible and large repositories and have a well documented API to access these repositories. First, user profiles of all 143,853 registered users on Codeforces was obtained using the Codeforces API. Next, for each of these users, details of all submissions made by these users to Codeforces was fetched for this study. This included more than 72 million submissions made to Codeforces. Submissions made to rated contests during its specified duration, submissions made to problems for practice, submissions that were successfully accepted or not, and those made in all accepted languages by the judge (including C/C++, Java, Ruby, Rust, Python, PHP, Kotlin, etc.) were included for analysis. The data obtained was largely clean except for missing entries for features that were not relevant to this research.

Features describing each user submission are shown in Table II. To this set of features, two new features - Experience based on time (exp_time) and experience based on the count of rated contests participated by the user (exp_contests) - were calculated and added for each submission entry. *exp_time* was calculated in days as the difference between the current date and the date of registration of the user on Codeforces. *exp_contests* involved more steps. First, to classify contests as rated contest hosted by Codeforces, details of all contests had to be obtained. Next, *contestId* for all Codeforces' rated contests were shortlisted and used to filter user's submissions based on the *problem.contestId* field to only include submissions made to rated contests. Next, using the *author.participantType* field, only submissions made by the user to the contest as a *Participant* were filtered. From this truncated submission list for each user, his *exp_contests* was then obtained.

Next, the nominal variable, *verdict*, in each submission, was converted to a numeric variable $0 \leq score \leq 1$ based on the verdict obtained against the test suite (*testset*) , where $score = 0$ implies that no submission has been made and

TABLE I. OJ RESPONSES TO USER SUBMISSIONS BASED ON QUALITY SUB-CHARACTERISTICS

OJ Verdict	Status of submission's execution	A	B	C
Wrong Answer (*WA*)	wrong output or some requirements not satisfied	X	X	-
Time Limit Exceeded (*TLE*)	exceeded maximal specified processing time limit	-	-	X
Memory Limit Exceeded (*MLE*)	exceeded specified RAM utilization limit (stack or heap)	-	-	X
Runtime Error (*RTE*)	runtime error occurred during execution	-	-	X
Accepted (*AC*)	passed all tests without exceeding resource or time limits	Y	Y	Y

Y - Satisfied, X - Not satisfied A-correct, B-complete, C-meets all specified performance and resource constraints

TABLE II. FEATURES DESCRIBING EACH SUBMISSIONS

#	Feature name	Datatype	#	Feature name	Datatype	#	Feature name	Datatype
1	id	int64	10	timeConsumedMillis	int64	19	author.members	object
2	user	object	11	memoryConsumedBytes	int64	20	author.participantType	object
3	contestId	int64	12	points	float64	21	author.ghost	bool
4	creationTimeSeconds	int64	13	problem.contestId	int64	22	author.startTimeSeconds	float64
5	relativeTimeSeconds	int64	14	problem.index	object	23	problem.points	float64
6	programmingLanguage	object	15	problem.name	object	24	problem.rating	float64
7	verdict	object	16	problem.type	object	25	author.room	float64
8	testset	object	17	problem.tags	object	26	author.teamId	float64
9	passedTestCount	int64	18	author.contestId	int64	27	author.teamName	object

$score = 1$ indicates an *AC*. Concept of *score*, and calculations related to *score* have been elaborated in [27].

A. ML Model Setup

In this experiment, we trained and tested seven ML regression models using traditional developer stats directly provided by OJs and use it to predict the *score* of developer in an unseen contest in a problem (E) at difficulty level $\theta = 5$ which was used as the target variable, Y. Traditional stats used as input features to the ML regression models were as follows: X = [*'rank'*, *'maxRank'*, *'rating'*, *'maxRating'*, *'position'*, *'exp_years*', '*exp_contests*'], The statistical distribution of features in X are given in Table III.

We decided to use symbolic ML representations in this experiments to better assess and study the explainability of a regression model as using subsymbolic models such as Deep Neural Networks or ensembles of simpler regression models will require the deciphering and subsequent understanding of the underlying complex black-box models which still remains a challenge [2].

The seven ML regression models were successfully trained and tested based on data of 143,853 registered users. The RMSE values and the prediction accuracy obtained using the seven ML regression models are given in Table IV. As the purpose of this research was not primarily to improve the prediction of a regression model by tuning the hyperparameters of the model, but to study the effect of features in improving the explainability of a model, we decided to use just one among the seven models. Among the seven ML regression models used, the *GradientBoost* regressor provided the least *RMSE* and the best *score* in our first experiments and was, thus, retained as the model of choice in this research.

The next section explains the evaluation of LIME and SHAP [28], two popular model-agnostic XAI tools that can be easily applied to comprehend ML models based on symbolic representations.

TABLE III. STATISTICS OF TRADITIONAL STATS

Feature	mean	std	min	max
rating	1074.4	454.9	-53.0	3757.0
maxRating	1130.3	473.7	227.0	3979.0
rank	1.0	1.8	0.0	10.0
maxRank	1.2	1.9	0.0	10.0
position	71927.0	41526.9	1.0	143853.0
exp_time	773.2	755.2	4.0	5234.0
exp_contests	12.6	23.7	0.0	816.0

TABLE IV. MODEL PREDICTION USING TRADITIONAL DEVELOPER STATS

ML Model used	RMSE values	accuracy
Elastic	0.1409	0.3106
Gradient Boost	0.1367	0.3507
Lasso	0.1408	0.3118
Linear	0.1407	0.3129
Random Forest	0.1442	0.2780
Ridge	0.1407	0.3129
XGBoost	0.1436	0.2837

B. LIME Analysis

LIME explains a prediction of a machine learning model, in our case the GradientBoost Regressor model, for a query point by finding important predictors and fitting a simple interpretable model. We applied LIME on our model and tested its behavior with ten randomly chosen query points, specifying five as the number of most important predictors to report. Based on these parameters, LIME was applied to our testset; LIME internally generated a synthetic data set, fitted a simple interpretable model of important predictors to it, and then used it to explain the predictions around the specified externally supplied ten selected query points. The explanations generated by LIME for the 10 query points and for our ML regression model are given in Table V.

It can be seen that *position* is consistently being used in all 10 explanations as the feature contributing the most to the prediction. It appears that LIME has identified four crisp categories of positions which are then used in the explanation. The second most important feature appears to be *maxRating* which also appears to have been divided into four separate

TABLE V. LIME EXPLAINABILITY

	Feature contribution 1	Feature contribution 2	Feature contribution 3	Feature contribution 4	Feature contribution 5
1	60871.00 ¡ position ¡= 93407.50	929.00 ¡ maxRating ¡= 1184.00	exp_contests ¿ 17.00	rank ¡= 0.00	870.00 ¡ rating ¡= 1107.00
2	position ¿ 93407.50	maxRating ¡= 929.00	maxRank ¡= 0.00	exp_contests ¡= 2.00	exp_time ¡= 295.00
3	30101.50 ¡ position ¡= 60871.00	1184.00 ¡ maxRating ¡= 1480.50	exp_time ¿ 1139.00	1107.00 ¡ rating ¡= 1422.00	0.00 ¡ maxRank ¡= 3.00
4	30101.50 ¡ position ¡= 60871.00	1184.00 ¡ maxRating ¡= 1480.50	exp_time ¿ 1139.00	0.00 ¡ maxRank ¡= 3.00	0.00 ¡ rank ¡= 3.00
5	position ¡= 30101.50	maxRating ¿ 1480.50	rating ¿ 1422.00	exp_contests ¿ 17.00	exp_time ¿ 1139.00
6	30101.50 ¡ position ¡= 60871.00	exp_time ¿ 1139.00	1184.00 ¡ maxRating ¡= 1480.50	1107.00 ¡ rating ¡= 1422.00	0.00 ¡ rank ¡= 3.00
7	60871.00 ¡ position ¡= 93407.50	929.00 ¡ maxRating ¡= 1184.00	2.00 ¡ exp_contests ¡= 6.00	maxRank ¡= 0.00	870.00 ¡ rating ¡= 1107.00
8	30101.50 ¡ position ¡= 60871.00	1184.00 ¡ maxRating ¡= 1480.50	295.00 ¡ exp_time ¡= 604.00	6.00 ¡ exp_contests ¡= 17.00	1107.00 ¡ rating ¡= 1422.00
9	60871.00 ¡ position ¡= 93407.50	929.00 ¡ maxRating ¡= 1184.00	exp_contests ¿ 17.00	295.00 ¡ exp_time ¡= 604.00	870.00 ¡ rating ¡= 1107.00
10	position ¿ 93407.50	929.00 ¡ maxRating ¡= 1184.00	maxRank ¡= 0.00	rating ¡= 870.00	2.00 ¡ exp_contests ¡= 6.00

categories and are used in the explanation. The exception to this is the 6th datapoint where LIME uses *exp_time* to provide the explanation. For the third and next important features there is a variation between the importance of features in prediction. For example, an XAI output generated in Natural Language to explain the first point may be as follows:

"*Since the user was placed between positions 60871 and 93497.50, and had a rating between 929 and 1184, and had participated in more than 17 contests and ranked at the lowest level, and had a rating between 870 and 1107, ...*"

Similarly, for the second point the explanation would have read as follows:

Since the user was placed at a position higher than 93497.50, and had a rating lower than or equal to 929, and had always been ranked at the lowest level, and participated in no more than 2 rated contests, and made his debut on Codeforces 295 days back , ...

Although this is easily understandable from the domain expert point of view, such explanations involving wide and deep decision trees and which have mathematics at their core, may overwhelm a typical end user even when provided in Natural Language.

To identify the feature importance in explanations, we executed LIME on 3000 separate data points and collated the results as shown in Table VI. Each row in the table shows details of one the six traditional stats, as specified below the table. The columns specify the feature importance. For example, *rating* (Feature 4) is the most important feature used in explaining 2925 of the 3000 points (97.5%) and is used as the second most important feature used to explain 75 of the total 3000 points (2.5%). Another important feature is *maxRating* (Feature 1) which was used as the most important feature to explain 75 of the 3000 data points (2.5%) and the second most important feature in explaining 2558 of the 3000 (85.26%) points. However, we see that the numbers are widely distributed among other features in the table cells. Such distribution of numbers in the table implies that the decision tree used in the explanation by LIME can get very wide clouding the explanation in NL when explaining multiple data points.

A metric that could capture this information could be represented as a k-tuple where k is the number of features used. In this experiment we have used $k = 5$ so we have a 5-tuple as follows: $< 97.5, 85.26, 37.2, 24.97, 19.7 >$ where each number in the tuple indicates the percent contribution of the feature to the k-th position. In this example, we have

TABLE VI. FEATURE IMPORTANCE USING LIME

	1	**2**	**3**	**4**	**5**	%
Feature 4	**2925**	75	0	0	0	20.15
Feature 1	75	**2558**	285	52	26	20.12
Feature 0	0	105	**1116**	762	449	16.33
Feature 5	0	134	650	**749**	591	14.26
Feature 6	0	134	650	749	**591**	14.26
Feature 3	0	4	132	359	695	7.99
Feature 2	0	1	84	296	645	6.89

features = [rating, maxRating, rank, maxRank, position, exp_time, exp_contests]

Feature 4 being used 97.5% of the time as the first feature in explaining a data point and Feature 1 being used 85.26% of the time as the second feature in explaining a data point. In an ideal case (a fully and consistently explainable model), each number in the k-tuple should be 100.

An ideal matrix for XAI would be a left diagonal sparse matrix of size $k \times k$, where k is the parameter to LIME specifying the number of features to include in the explanation.

C. SHAP Analysis

SHAP, like LIME, is model-agnostic and can be applied to explain models based on symbolic representations. It stands for SHapley Additive exPlanations and attempts to explain the prediction of an instance by computing the contribution of each feature to the prediction. We applied SHAP to our regression model and to evaluate its explainability on our dataset. Fig. 2 shows the SHAP plot for the seven traditional stats (listed below the figure). Similar to LIME, SHAP also ranked *position* as its most important feature followed by *maxRating*. But we can see that *exp_contests*, *rating*, and *exp_time* are also highly used.

Having a plot shape which is top-heavy, like an exaggerated overgrown mushroom, helps improve the explainability of an XAI tool. Carefully assembling and curating a good dataset can bring such a shape to the plot and increase its explainability.

V. ADDING DOMAIN KNOWLEDGE TO DATASET

Based on lessons learnt from the previous experiment, we decided to incorporate some domain knowledge directly into the dataset. This first involved basic feature engineering to identify and eliminate unimportant features. Next, having studied different OJ environments, their rated contests and user behavior during contests over the past several years, we decided to include features that implicitly incorporated domain knowledge into the dataset. The next subsections give a few illustrative examples of domain knowledge missing from the

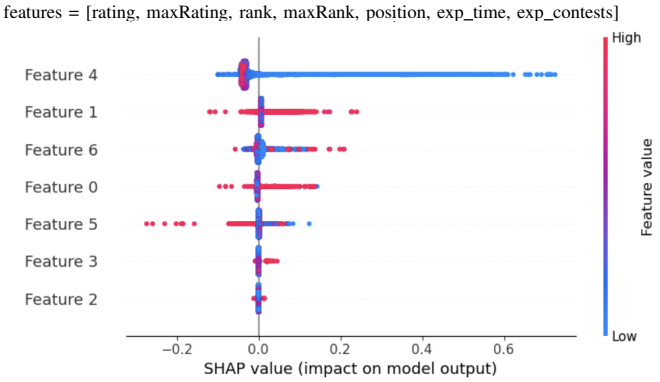


Fig. 2. SHAP Plot showing feature importance for traditional stats.

previous dataset because of which both the predictability of the model and the XAI aspect of the model were affected. This involved introducing new features, deriving new features from existing features, identifying outliers based on these newly generated features, classifying and segregating classes of submissions, we created a new dataset comprising of more than 40 new features in addition to the previous seven traditional stats used in the previous experiment. By including this domain knowledge into the dataset as additional features, and appropriately segregating and curating data, and training the ML regression model on this dataset, we show that the predictability and explainability of the trained model can be substantially increased.

A. Explaining Rank, Rating and Positional Related Inconsistencies

Rank, *rating* and *position* on an OJ may not give a complete picture of the ability and skills of a developer. For example, a low *rank* and *rating* on the OJ does not necessarily imply that the developer has poor programming ability. This is because every new user registering on an OJ, is assigned a default low *rank*, *rating* and *position* regardless of his programming experience or capability. It will take regular participation in rated contests on the OJ and consistently good performances in them for a developer to rise up the leaderboard and to be assigned a high *rank*, *rating* or *position* by the OJ. The rating algorithm used by Codeforces is a variation of the Elo rating [29] used in sports. Understanding the Elo rating system will help explain why an old user on the OJ with good *rating* can fall down the leaderboard with a few bad consecutive performances in contests. This knowledge is not explicitly built into the previous dataset. Identifying and explaining such instances will be challenging to an XAI tool in the existing case. For example, how could an XAI tool explain the situation of a newly registered user on the OJ consistently and successfully solving the difficult problems in contests?

1) Correction mechanism: By associating user submission history to user performances, could help generate better and more reasonable XAI outputs.

B. Upsets *in Contests*

Similar to contests in sports, seeing upsets happen in rated contests is not uncommon. As illustration of upsets, consider the two examples given in Fig. 4. The first example shows the five most recent performances of an average programmer. We can see that he had successfully solved only two easy problems while consistently attempting to solve the third one ($\theta = 3$). However, in the unseen contest he was able to successfully solve the first four problems ($score = 1$) and obtain a high score in problem E (at $\theta = 5$). The second example, shows the performances of a good programmer in the last five most recent contests.We can see that he consistently obtained high scores in all the five problems. But in the unseen contest, however, he performs very badly, not able to successfully solve even the two most easy problems. How could such datapoints be presented by XAI tools?

1) A possible XAI explanation in natural language: "This average user has caused an upset by solving the most difficult problems in the contest."

2) Correction mechanism: These *upsets* or erratic and temperamental performances [27] by contestants can be seen as outliers or stray occurrences in the dataset. Correctly identifying and labelling such *upsets* in the dataset or eliminating them altogether will help improve the accuracy of the model and increase the explainability aspect of the model.

C. Divisions: Different Playing Fields

The *problem.index* field in the *submissions* table (Feature no. 14 in Table II) indicates the problem index (A..E) and hence specifies the problem's difficulty level ($1 \leq \theta \leq 5$). Closer inspection of the dataset reveals a strange trend - higher ranked and rated contestants have difficulty in successfully solving problems D and E, while many lower ranked and rated contestants regularly seem to be successfully solving problem E. Domain knowledge says that all E's are not equal. Users on Codeforces are classified into four divisions based on their rating. Higher rated contestants play (participate) in the highest *Division* (*Division* 1). Rated contests are held separately for each *Division*. While users from higher *Divisions* can register and participate in contests targeted for lower *Divisions*, they are not counted towards their rating. Problems in all rated contests, regardless of *Division*, are indexed from *A* to *E*. This implies that a problem indexed as *E* in a contest at a lower *Division* could be presented as index *A* in a contest at a higher *Division*. This muddles the concept of problem difficulty (θ) and brings inconsistency in the predictability of the regression model. How could such datapoints be presented by XAI tools?

1) A possible XAI explanation in natural language: "This user is from a lower Division and thus could solve problem E which, in reality, may not be too difficult to solve"

2) Correction mechanism: Classifying and segregating submissions based on contestant *rating* into Divisions and possibly training and testing separate ML regression models based on Divisions. This mechanism will ensure the XAI outputs are more natural and understandable.

D. Oscillations Between Divisions

Oscillations between Divisions is an extension of the previous explanation of Divisions. Consider a developer whose current rating is close to the boundary of the next higher Division. A good performance in a rated contest in his existing

(a) Good programmer in good health (b) Good programmer in poor health

(c) Average programmer in good health (d) Average programmer in poor health

(e) A healthy programmer showing consistently good growth (f) A programmer showing a chronic health condition

Fig. 3. Sample health cards showing developer health over a 7-contest window ($w = 7$).

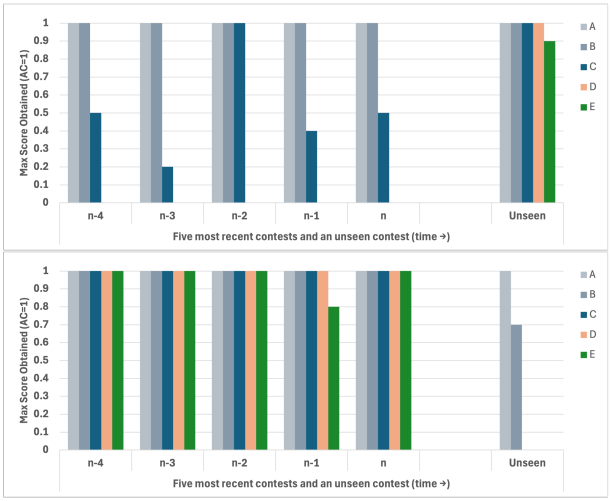


Fig. 4. Upsets in contests.

Division will increase his rating and will subsequently place him in the next higher Division. Similarly, a poor performance by a user in a contest whose rating is close to the boundary of the lower Division, may be placed in the lower Division after the contest. A situation may occur when the ratings fluctuate at the Division boundary causing a user to consistently solve problems D and E when placed in the lower division, and afterward fail to solve problems D and E when placed in the higher Division and vice versa. We refer to this as phenomenon as *oscillations between Divisions*. This common but strange behavior negatively impacts the prediction of a ML model and can be difficult to explain by an XAI tool. How could such datapoints be presented to end user of an XAI tool?

1) A possible XAI explanation in natural language: "*This user's currently rating is at the Division boundary and he is, thus, demonstrating this oscillating behavior.*"

2) Correction mechanism: Specifying a boundary threshold (δ) and segregating users from the dataset whose rating falls within this threshold ($\pm\delta$) at the Division boundary. This mechanism will ensure the XAI outputs are more natural and understandable.

E. Joint Contests

OJs frequently organize joint contests between two different Divisions with problem being identified using the same problem index. However, the rating change calculations differ between divisions. This is not very explicit in the dataset and such contests will have to be identified and marked as such in the dataset.

1) A possible XAI explanation in natural language: "*This user's participation is probably in a joint contest between Divisions and thus has not been able to solve problem E in the higher Division level.*"

2) Correction mechanism: Manually identifying and tagging joint contests as such, will help generate better XAI outputs which are more natural and understandable.

F. Health Card and Development History

Similar to clinical history of a patient [16] and player stats which show past performances and record of a soccer player, a development history could record the history of a developer. This history could provide various insights into a developer's behavior over time. This could be semantically arranged and structured to provide a valuable source of information for XAI applications . Unfortunately, there is currently no standard way of documenting this history - past achievements related to code development, role and contribution to open-source projects, awards obtained in hackathons and programing contests, etc. remain disparate entities and connected together only on the CV of a developer if a job change is desired or to a social media page.

A performance in a single recruitment contest or a hackathon cannot be used as the sole criteria by HR to recruit a candidate. There may be cases of false positive or false negatives.

1) A bad programmer being classified as good - An average candidate may perform extremely well in such a contest if he is lucky enough to get a problem previously 'seen' by him or that he has worked on
2) A good programmer being classified as bad - A good programmer who may not perform well in such a contest because of health-related or personal reasons on that particular day.

However, a record of the developer's recent history augmented with the performance in the recruitment contest can aid HR in better decision making.

1) Reflection of developer health: A health card of a developer provides an clear insight into the health, maturity level and growth of a developer. Fig. 3 shows examples of five health cards. Each health card shows a windows into scores obtained by contestants in their last seven rated contests ($w = 7$) with the rightmost contest being the most recent contest that the user participated in. Each contest has five problems (A..E) at varying difficulty levels (θ) with A being at the lowest level of difficulty ($\theta = 1$) and E at relatively the highest level of difficulty ($\theta = 5$) among the five problems. Good developers are identified by their capability of successfully and consistently solving all five problems with an AC ($score \sim 1$). Average developers are able to solve the first few problems ($\theta \leq 3$) consistently with a $score = 1$, but have difficulty successfully solving problems at the higher level of difficulty ($\theta \geq 4$) and usually end up obtaining $0 \leq score < 0.7$ for these problems.

As can be seen from Fig. 3(a), the first user appears to be good developer as demonstrated by his consistently and (almost) successfully solving all five problems in all his seven past contests. His *rank*, *position* and *rating* on the OJ will continue to remain stable with minor fluctuations. On the other hand, the health card of the second user shown in Fig. 3(b) shows a different trend. The user appears to have a history of solving all five problems in contests. However, his performance in the more recent contests seems to have deteriorated. We refer to this performance deterioration as a sign of *poor health* of the developer. The term *poor health* is not absolute but relative to a developers's past performances. In this example, the user has successfully solved a few problems. But based on his demonstrated capability of solving all problems in the past, we say that his health is *poor*. Because of his poor recent performances, his *rank*, *position* and *rating* will take a beating and fall to levels that may not reflect the developer's actual capability.

Likewise, we have the health cards of two average programmers. Fig. 3(c) shows the health card of a healthy average programmer. We label him *healthy* as his health card shows a consistent performance with the developer consistently solving the first few problems in the most recent seven contests. The *rank*, *position* and *rating* of this developer on the OJ will continue to remain stable with minor fluctuations. On the other hand, the average programmer's health card in Fig. 3(d) shows a deteriorating performance - although he had a history of demonstrated capability in solving problems A and B in a few recent contests, he has consistently failed to successfully solve these problems in the more recent contests. Similar to the effect of deteriorating health of a good developer, the *rank*, *position* and *rating* on this developer on the OJ will drop to levels that do not rightly indicate the capability of this developer.

This *poor health* that we refer to in this paper is similar to the *loss of form* of a player referred to in sports. There is a high probability that a developer in poor health may not be able to perform his best in a recruitment hackathon or a contest much like an *out-of-form* player's performance in a game because of an injury, illness or some personal issue. This developer, as such, should not be evaluated by a single contest in his current poor form.

The health card in Fig. 3(e) shows a developer in good health and demonstrating good and steady growth. We see that his capability as a developer has matured over the last few contests. A few contests back, he was only able to solve problems at $\theta \leq 2$ but has grown in his capability as a developer to successfully solve problems at $\theta \leq 4$ while able to attempt problems at 5.

These health cards in the figure show contests over a window size of $w = 7$. Larger window sizes will give more meaningful and deeper insights into the growth, consistency and maturity of a programmer over a longer time period.

2) Structure of a developer's health card: A health card reflects the performance of a user in each of the rated contests that he has attempted on the OJ. A contest is defined as $C = \{c_1, c_2, ..., c_n\}$ where each c_i gives a summary of the performance of the developer in the contest *i*. The size of the set, $|C|$, gives the contest experience or the number of rated contests the developer has participated in on the OJ. For the scope of this paper, each contest has five problems (A..E) at increasing level of difficulty, with A being at the easiest level ($\theta = 1$) and E at the most difficult level ($\theta = 5$).

Each c in turn consists of sets of 5-tuples represented formally as $c = \{s_1, s_2, ...s_m\}$. where each s is a 5-tuple representing values associated with the five problems (A..E) in the contest. The size of the *set, m*, is contextual and can be varied based on the user need. For example, s_1 could be used to

represent the final verdict awarded by the OJ for the contest where verdicts are assigned representative scores between 0 and 1.0 where 0 indicates that the user has made no submission to the problem in the contest and 1.0 represents an AC with other verdicts (WA, RTE, TLE, MLE, etc.) falling within this range. Similarly, s_2 could be used to record attempts made for each problem in a contest. Other stats that could be stored are numbers of WAs for each problem, number of TLEs obtained for each problem, etc.

For example, consider the following representation of data of a programmer for contest *j*.

$$c_j = \{score_j, attempts_j, wa_j, tle_j, rte_j ...\}$$

where each $c_j \epsilon C$ represents a rated contest that a user has participated in on the OJ. It could contain various details about the performance of the user in problems in the contest at difficulty levels $1 \leq \theta \leq 5$ as 5-tuples. Some of the important statistics that could be included in c_j are given below. *score* is a 5-tuple that gives the maximum score that a user has been able to attain among all attempts that were submitted to the OJ during the contest. Consider the example below:

$$score_j = (1.0, 1.0, 1.0, 1.0, 0)$$

In this contest, the user has been able to successfully solve the first four problems (A..D) but did not solve problem E. The 5-tuple *score* does not indicate the number of attempts (shots) that were taken. The 5-tuple *attempts* indicates the total number of submissions, both successful and unsuccessful, that the user made to the OJ during the contest. Consider the example below

$$attempts_j = (1.0, 1.0, 1.0, 8.0, 0)$$

The user made one submission each to problem A..C, made eight attempts to solve problem D and no attempts at solving problem E. This 5-tuple gives an indication to the ability of a developer to understand the given problem statement. The verdicts returned by the OJ for each of the submissions and for each of the problems are captured in their respective 5-tuples. Consider the 5-tuple for wa_j below

$$wa_j = (1.0, 1.0, 1.0, 1.0, 0)$$

This indicates that the user has made one wrong submission for each of the problems A..D. The 0 for problem could mean two things - either he obtained no WA for problem E for any of the submissions made or he obtained no WA because he made no submission to problem E. This can be better understood from $attempts_j$. Similarly, we have the 5-tuples tle_j and rte_j, which give an indication to the submissions which were awarded TLE and RTE for each of the five problems.

For contestants who have participated in many contests, the size of set C can get large. Also, if we decide to store a large set of data for each contest, the size of the encoding vector can become fairly large. Although this could be seen as a limitation, it adds more explainability to the model. However, this limitation can be overcome by constraining the size of set *C*. The set *C* could be seen in two forms:

1) The set *C* contains the complete history of the developer on the OJ where $|C|$ gives the total sum of rated contests that the developer has participated in. Although this set can get large, this provides insights into the growth and maturity of a developer since his registration on the OJ.
2) The set *C* contains only a window to the most recent *w* contests and provides a glimpse to the health or most recent form of a developer. This can be see in Fig. 3 (Health chart) with where $w = 7$.

[Explain the significance of the two sets] EAI can add transparency and a clearer explanation and justification to the decision making process by presenting the health chart of the developer in a visually intuitive form while additionally providing the set *C*. This will also help in increasing trust in the system.

VI. Enhancing XAI with Derived Data

The main objective of this research was to bring more explainability to the ML models by focusing more on the data aspect rather than just enhancing the performance of ML models though hyper-parameter tuning. This was achieved by enhancing data quality through carefully creating and adding new features to the dataset and by creatively mitigating the effect of noise and outliers using knowledge of the domain.

One of the major changes made to the dataset used in the previous experiment was to incorporate developer history. This proved to be a double-edged sword - it increased the performance of the model while providing more meaningful data to enhance the explainability of the XAI tools. The resultant dataset considerably improved the r2-score of the GradientRegressor model from 0.3507 to 0.9834 (63.27%) which was a substantial improvement.

Additionally, the health chart of a user can be easily extracted from the dataset to add more human-understandable explanation to the XAI generated output.

A. LIME - Results and Discussion

Similar to the previous experiment, we applied LIME to the GradientBoost Regressor model trained on the enhanced dataset and tested its behavior with ten randomly chosen query points, specifying five as the number of most important predictors to report. For these parameters, LIME internally generated a synthetic data set, fitted a simple interpretable model of important predictors to it, and then used it to explain the predictions around the specified externally supplied ten selected query points. The explanations generated by LIME for the 10 query points and for our ML regression model in this experiment are given in Table VII.

Compared to the explanations in the previous experiment, we note that one feature is consistently and dominantly used in all 10 explanations as the feature contributing the most to the prediction and using the same decision. This is reflected in the second and third most important features used in the explanation of the query points with no exceptions. When

generating an explanation for this in Natural Language, the XAI output would be consistent across most, if not all, data points. For example, explanations for the first two data points could read as as follows:

1) "*Since the user could not successfully solve problems B, E and C even though he made feeble attempts to solving problems E and C in the last five contests he participated in .....*"
2) "*Since the user could not successfully solve problems B, E, C and A and made no attempt to solve problem D in the last five contests he participated in.....*"

Compared to the possible explanations that could be generated in the previous experiments, these are more understandable and consistently presented to the end user. A caveat - in the dataset for Codeforces, and generally in most OJs, a large majority of registered participants are not able to successfully solve most problems. This explains the apparent bias visible in the choice of query points in Table VII.

To further identify the feature importance in explanations, we executed LIME on 3000 separate data points and collated the results as shown in Table VIII.

The table has been limited to show just twenty of the total features used in this experiment. Each row in the table shows details of one feature. The columns specify the feature importance. In this experiment, Feature 16 was the most important feature used in explaining all of the 3000 points (100%). The second and thirst most important features were Feature 19 and Feature 17, respectively, which were used to explain all of the 3000 data points (100%). However, we see that the numbers are evenly distributed among other four features in the table cells for the fourth and fifth most important feature. This distribution of numbers in the table differs from the numbers seen in the previous experiment. We can be infer from the features and the range selected that the decision trees used in the explanations may not be very wide which will aid the visual explanation generated by LIME when explaining multiple data points.

In this experiment 5-tuple used to capture the explainability of the model read as follows: $< 100, 100, 99.13, 24.8, 21.8 >$. This is a substantial improvement over the previous 5-tuple which was $< 97.5, 85.26, 37.2, 24.97, 19.7 >$. Comparing the two, we can observe that each feature contribution to the explanation has improved.

Also, and as pointed out earlier, an ideal matrix for XAI would be a left diagonal sparse matrix of size $k \times k$, where k is the parameter to LIME specifying the number of features to include in the explanation. Wee see that this matrix is a left diagonal matrix at $k = 3$ which was not the case in the previous experiment.

B. SHAP: Results and Discussion

We applied SHAP to evaluate the GradientBoost Regressor model trained on the enhanced dataset for explainability. Fig. 5 show the SHAP plot for twenty of the total features present in the dataset. Similar to LIME, SHAP also ranked Feature 16 as the most important feature followed by Feature 19 followed by Features 17, 18 and 15.

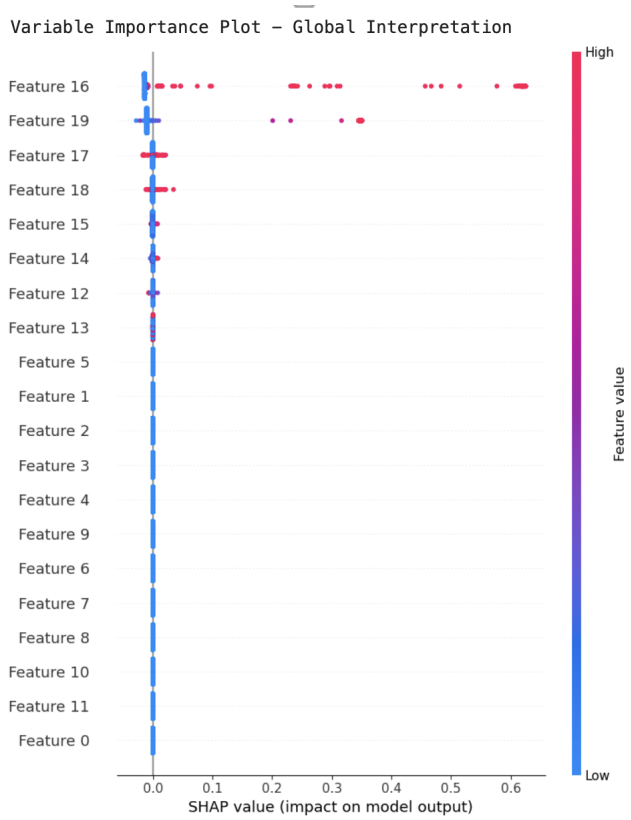


Fig. 5. SHAP values for ML model using the enhanced dataset.

As pointed out earlier, and ideal SHAP plot shape would be a top-heavy one, like an exaggerated overgrown mushroom. The SHAP plot obtained in this experiment appeared to be just that - this greatly aids in improving the explainability aspect of the XAI tool. This was a result of carefully assembling and curating the dataset to bring such a shape to the SHAP plot and, subsequently, to increase its explainability.

VII. Conclusion

Bringing more transparency to the decision making process in fields deploying ML tools is important in various fields. This implies that ML tools need to be designed in such a way that they are more understandable and explainable to the end users while also increasing trust in them. The field of XAI, although a mature area of research, is increasingly being seen as a solution to address these missing aspects of ML systems.

The focus of this work was on improving the transparency of the decision making process in recruitment of software developers using Online Judges. As the field of software development attracts talent at various levels for the high-paying lucrative jobs that is has, it is important to correctly identify and differentiate highly skilled developers from developers who are adept at only performing regular and mundane programming jobs. Also, HR recruiting agents needs to report back to their managers and justify why certain candidates were selected and why some were rejected.

To address this, we built a regressor model that can help differentiate developers based on their ability, while identifying

TABLE VII. LIME EXPLAINABILITY

	Feature contribution 1	Feature contribution 2	Feature contribution 3	Feature contribution 4	Feature contribution 5
1	B_solved_5 ¡= 0.00	E_solved_5 ¡= 0.00	C_solved_5 ¡= 0.00	E_attempts_5 ¿ 1.33	C_attempts_5 ¿ 0.50
2	B_solved_5 ¡= 0.00	E_solved_5 ¡= 0.00	C_solved_5 ¡= 0.00	D_attempts_5 ¡= 0.00	A_solved_5 ¡= 0.00
3	B_solved_5 ¡= 0.00	E_solved_5 ¡= 0.00	C_solved_5 ¡= 0.00	A_solved_5 ¡= 0.00	D_attempts_5 ¡= 0.00
4	B_solved_5 ¡= 0.00	E_solved_5 ¡= 0.00	C_solved_5 ¡= 0.00	A_solved_5 ¡= 0.00	E_attempts_5 ¡= 0.00
5	B_solved_5 ¡= 0.00	E_solved_5 ¡= 0.00	C_solved_5 ¡= 0.00	D_solved_5 ¡= 0.00	E_attempts_5 ¡= 0.00
6	B_solved_5 ¡= 0.00	E_solved_5 ¡= 0.00	C_solved_5 ¡= 0.00	0.50 ¡ E_attempts_5 ¡= 1.33	0.00 ¡ C_attempts_5 ¡= 0.10
7	B_solved_5 ¡= 0.00	E_solved_5 ¡= 0.00	C_solved_5 ¡= 0.00	A_solved_5 ¡= 0.00	E_attempts_5 ¡= 0.00
8	B_solved_5 ¡= 0.00	E_solved_5 ¡= 0.00	C_solved_5 ¡= 0.00	C_attempts_5 ¡= 0.00	E_attempts_5 ¡= 0.00
9	B_solved_5 ¡= 0.00	E_solved_5 ¡= 0.00	C_solved_5 ¡= 0.00	C_attempts_5 ¡= 0.00	E_attempts_5 ¡= 0.00
10	B_solved_5 ¡= 0.00	E_solved_5 ¡= 0.00	C_solved_5 ¡= 0.00	D_solved_5 ¡= 0.00	0.10 ¡ D_attempts_5 ¡= 1.00

TABLE VIII. LIME VALUES FOR DERIVED FEATURES

	1	**2**	**3**	**4**	**5**	**%**
Feature 16	**3000**	0	0	0	0	20
Feature 19	0	**3000**	0	0	0	20
Feature 17	0	0	**2974**	19	2	19.97
Feature 18	0	0	6	**744**	365	7.43
Feature 14	0	0	6	655	**654**	8.77
Feature 12	0	0	5	571	630	8.04
Feature 15	0	0	7	536	643	7.91
Feature 13	0	0	2	475	706	7.89
Feature 0	0	0	0	0	0	0
Feature 1	0	0	0	0	0	0
Feature 2	0	0	0	0	0	0
Feature 3	0	0	0	0	0	0
Feature 4	0	0	0	0	0	0
Feature 5	0	0	0	0	0	0
Feature 6	0	0	0	0	0	0
Feature 7	0	0	0	0	0	0
Feature 8	0	0	0	0	0	0
Feature 9	0	0	0	0	0	0
Feature 10	0	0	0	0	0	0
Feature 11	0	0	0	0	0	0

and ignoring their erratic (temperamental) performances during contests. We showed how both the ML model and the underlying dataset used in training and testing the model can impact the explainability of the model. The underlying dataset that was readily available from the OJ, was enhanced by adding more features and creating new derived features based on our knowledge of the domain. This was done to add more explainability to the model and to increase its predictability and performance accuracy. We also showed how XAI can be actively and iteratively used during pre-deployment to improve the quality of the dataset and to improve the prediction accuracy of the regression model. These iterative changes helped improve the r2-score of our GradientRegressor model from 0.3507 to 0.9834 (63.27%) which was a substantial improvement. We also showed how the consistency and explainability of LIME and SHAP, the XAI tools used in this research, increased over the iterations.

We believe that the work presented in this paper, has great applicability to areas other than developer recruitment. For example, it could be used by project managers to suggest focussed training regimes for developers in their team, to recruit developers for specialized domains, by coaches at Universities to better select their programming team, and in academia where students' performance in programming courses need be predicted to take early remedial action.

REFERENCES

[1] J. Lötsch, D. Kringel, and A. Ultsch, "Explainable artificial intelligence (xai) in biomedicine: Making ai decisions trustworthy for physicians and patients," *BioMedInformatics*, vol. 2, no. 1, pp. 1–17, 2021.

[2] A. Adadi and M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence (xai)," *IEEE access*, vol. 6, pp. 52138–52160, 2018.

[3] A. Das and P. Rad, "Opportunities and challenges in explainable artificial intelligence (xai): A survey," *arXiv preprint arXiv:2006.11371*, 2020.

[4] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, *et al.*, "Explainable ai (xai): Core ideas, techniques, and solutions," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–33, 2023.

[5] M. Pocevičiūtė, G. Eilertsen, and C. Lundström, "Survey of xai in digital pathology," *Artificial intelligence and machine learning for digital pathology: state-of-the-art and future challenges*, pp. 56–88, 2020.

[6] G. Marín Díaz, J. J. Galán Hernández, and J. L. Galdón Salvador, "Analyzing employee attrition using explainable ai for strategic hr decision-making," *Mathematics*, vol. 11, no. 22, p. 4677, 2023.

[7] D. Hangartner, D. Kopp, and M. Siegenthaler, "Monitoring hiring discrimination through online recruitment platforms," *Nature*, vol. 589, no. 7843, pp. 572–576, 2021.

[8] G. Guizzardi and N. Guarino, "Explanation, semantics, and ontology," *Data & Knowledge Engineering*, vol. 153, p. 102325, 2024.

[9] A. Ru and F. Khosmood, "Hackathons for workforce development: A case study," in *Proceedings of the 5th International Conference on Game Jams, Hackathons and Game Creation Events*, pp. 30–33, 2020.

[10] S. Wasik, M. Antczak, J. Badura, A. Laskowski, and T. Sternal, "A survey on online judge systems and their applications," *ACM Computing Surveys (CSUR)*, vol. 51, no. 1, pp. 1–34, 2018.

[11] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, *et al.*, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information fusion*, vol. 58, pp. 82–115, 2020.

[12] G. D. A. DW, "Darpa's explainable artificial intelligence program," *AI Mag*, vol. 40, no. 2, p. 44, 2019.

[13] D. Gunning and D. Aha, "Darpa's explainable artificial intelligence (xai) program," *AI magazine*, vol. 40, no. 2, pp. 44–58, 2019.

[14] V. Vimbi, N. Shaffi, and M. Mahmud, "Interpreting artificial intelligence models: a systematic review on the application of lime and shap in alzheimer's disease detection," *Brain Informatics*, vol. 11, no. 1, p. 10, 2024.

[15] K. Hauser, A. Kurz, S. Haggenmüller, R. C. Maron, C. von Kalle, J. S. Utikal, F. Meier, S. Hobelsberger, F. F. Gellrich, M. Sergon, *et al.*, "Explainable artificial intelligence in skin cancer recognition: A systematic review," *European Journal of Cancer*, vol. 167, pp. 54–69, 2022.

[16] C. Panigutti, A. Perotti, and D. Pedreschi, "Doctor xai: an ontology-based approach to black-box sequential data classification explanations," in *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pp. 629–639, 2020.

[17] A. Gramegna and P. Giudici, "Shap and lime: an evaluation of discriminative power in credit risk," *Frontiers in Artificial Intelligence*, vol. 4, p. 752558, 2021.

[18] H. Sheridan, D. O'Sullivan, and E. Murphy, "Ideating xai: an exploration of user's mental models of an ai-driven recruitment system using a design thinking approach," 2022.

[19] T. Clement, N. Kemmerzell, M. Abdelaal, and M. Amberg, "Xair: A systematic metareview of explainable ai (xai) aligned to the software development process," *Machine Learning and Knowledge Extraction*, vol. 5, no. 1, pp. 78–108, 2023.

[20] Z. Huang, H. Yu, G. Fan, Z. Shao, M. Li, and Y. Liang, "Aligning xai explanations with software developers' expectations: A case study with code smell prioritization," *Expert Systems with Applications*, vol. 238, p. 121640, 2024.

[21] M. A. Awal and C. K. Roy, "Evaluatexai: A framework to evaluate the reliability and consistency of rule-based xai techniques for software analytics tasks," *Journal of Systems and Software*, vol. 217, p. 112159, 2024.

[22] S. Roy, G. Laberge, B. Roy, F. Khomh, A. Nikanjam, and S. Mondal, "Why don't xai techniques agree? characterizing the disagreements between post-hoc explanations of defect predictions," in *2022 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, pp. 444–448, IEEE, 2022.

[23] M. Mahbubur Rahman, B. Chandra Das, A. A. Biswas, and M. Musfique Anwar, "Predicting participants' performance in programming contests using deep learning techniques," in *International Conference on Hybrid Intelligent Systems*, pp. 166–176, Springer, 2022.

[24] Codeforces, "https://codeforces.com."

[25] ICPC, "https://icpc.global."

[26] S. S. Skiena and M. A. Revilla, "Programming challenges: The programming contest training manual," *Acm SIGACT News*, vol. 34, no. 3, pp. 68–74, 2003.

[27] W. Ahmed and A. Harbaoui, "Is this code the best? or can it be further improved? developer stats to the rescue," *IEEE Access*, vol. 12, pp. 144395–144411, 2024.

[28] A. M. Salih, Z. Raisi-Estabragh, I. B. Galazzo, P. Radeva, S. E. Petersen, K. Lekadir, and G. Menegaz, "A perspective on explainable artificial intelligence methods: Shap and lime," *Advanced Intelligent Systems*, p. 2400304, 2024.

[29] Wikipedia, "Elo rating system," September 2024.

Secret Sharing as a Defense Mechanism for Ransomware in Cloud Storage Systems

Shuaib A Wadho[1], Sijjad Ali[2], Asma Ahmed A. Mohammed*[3], Aun Yichiet[4], Ming Lee Gan[5], Chen Kang Lee[6]
Faculty of Information and Communication Technology, Universiti of Tunku Abdul Rahman, Kampar, Malaysia[1,4,5,6]
College of Computer Science and Software Engineering, Shenzhen University, China[2]
Department of Computer Science, University of Tabuk, Tabuk, Saudi Arabia*[3]

***Abstract*—Ransomware is a prevalent and highly destructive type of malware that has increasingly targeted cloud storage systems, leading to significant data loss and financial damage. Conventional security mechanisms, such as firewalls, antivirus software, and backups, have proven inadequate in preventing ransomware attacks, highlighting the need for more robust solutions. This paper proposes the use of Secret Sharing Schemes (SSS) as a defense mechanism to safeguard cloud storage systems from ransomware threats. Secret sharing works by splitting data into several encrypted shares, which are stored across different locations. This ensures that even if some shares are compromised, the original data remains recoverable, providing both security and redundancy. We conducted a comprehensive review of existing secret sharing schemes and evaluated their suitability for cloud storage protection. Building on this analysis, we proposed a novel framework that integrates secret sharing with cloud storage systems to enhance their resilience against ransomware attacks. The framework was tested through simulations and theoretical evaluations, which demonstrated its effectiveness in preventing data loss, even in the event of partial compromise. Our findings show that secret sharing can significantly improve the reliability and security of cloud storage systems, minimizing the impact of ransomware by allowing data to be reconstructed without paying a ransom. The proposed solution also offers scalability and flexibility, making it adaptable to different cloud storage environments. This research provides a valuable contribution to the field of cloud security, offering a new layer of protection against the growing threat of ransomware.**

***Keywords*—Ransomware; secret sharing; cloud storage; data leakage; reliability**

I. Introduction

Cyber attacks in the form of ransomware have now become one of the most frequent and most vicious types of cybercrime [1]. These attacks usually involve software that secretly encrypts a person's files, and hence becomes inaccessible until a certain amount of money is paid to the attackers for the decryption key [2], [3]. The rise of ransomware is attributed to the fact that the attackers can demand and receive payments in cryptocurrencies and continued expansion of digitalization in different industries. WannaCry and NotPetya attacks are some of the most infamous ransomware attacks that have shown the extent of the damage that can be caused to critical infrastructure, healthcare, finance, and everyday users [4].

In our research article Cloud storage services have revolutionized data management by offering scalable, accessible, and cost-effective solutions for storing vast amounts of information. Organizations and individuals alike benefit from the convenience and flexibility provided by cloud storage. However, this dependence on cloud services also presents significant security challenges. As more sensitive and critical data is stored in the cloud, it becomes an attractive target for cybercriminals. The centralized nature of cloud storage makes it vulnerable to various attacks, including data breaches, insider threats, and ransomware.

A. Motivation

The more frequent and complex ransomware attacks on the cloud storage systems clearly show that protect measures are insufficient. The conventional security measures including antivirus, firewall and data backup may not be effective in dealing with the modern day hackers. Ransomware can spread through the system and encrypt data within a short time; thus, organizations have no choice but to pay the ransom or lose their data.

Thus, secret sharing schemes may be considered as a rather effective approach to increasing cloud storage security. These cryptographic techniques include: this is a method of dividing a secret, for instance, a piece of data into several parts and spread out. Some of these shares are adequate to reconstruct the original data and, thus, the system offers both security and redundancy. Thus, using secret sharing schemes in cloud storage, data can be protected from ransomware attacks since even if some shares are stolen, the original data is still restorable.

B. Objectives

The purpose of this paper is to analyze how secret sharing schemes can help in combating ransomware attacks in cloud storage systems, which are explained followings:

1) Review Existing Secret Sharing Schemes: It is an analysis of many types of schemes for secret sharing and assess how suitable they are for the cloud storage security.

2) Propose a Novel Framework: Propose a design of an architecture that incorporates secret sharing techniques with cloud storage to strengthen the protection against ransomware threats. This framework will describe the steps of partitioning the data, encrypting the data, distributing the data and recovery of the data.

3) Evaluate Effectiveness: Evaluate the preparedness of work and acquire robust information based on simulations and theoretical investigation of the proposed framework about how well it will perform and how easily it can be penetrated by ransomware attacks while not compromising significantly on system throughput.

C. Contributions

The contributions of this paper are threefold:

1) Comprehensive Review: In this paper, we present a detailed analysis of the SSSs, their relevance in cloud storage security, and the most appropriate ones for combating ransomware.

2) Novel Framework: This paper proposes a new approach that incorporates secret sharing techniques to fortify the security of cloud storage. This framework outlines the measures to follow in order to incorporate SSS with ordinary cloud storage systems and guidelines for the same.

3) Evaluation and Analysis: To assess the efficacy of the proposed framework we have carried out a number of simulation and theoretical exercises. Thus, the proposed framework can be considered as an effective means of countering ransomware threats in cloud storage systems.

D. Structure of the Paper

The remainder of this paper is organized as follows: Section II presents the related work in the areas of ransomware protection and secret sharing techniques. Section III presents the definitions. The system proposed in this paper is described in Section IV, which includes the architecture and the way the framework will be implemented. Sections V and VI, respectively explains the experiment setup and the results of the experiment which prove that the framework is useful and efficient. Section VII looks at the security, usability and the major constraints of the proposed framework. Section VIII presents the discussions. Lastly, Section IX offers the summary of the research and the recommendations for further studies.

Thus, this paper focuses on the problem of ransomware in the cloud storage context and offers a new idea to overcome this issue using the secret sharing technique. The following framework is presented as a viable way of mitigating the current and future risks of ransomware attacks and thus maintaining the security of data stored in the cloud services.

E. Preliminaries

We used mathematical symbols throughout the paper, which is explained below in Table I.

TABLE I. PRELIMINARIES: SYMBOLS AND DESCRIPTIONS

Symbols	Description
S	Secret to be shared
n	Total number of shares
t	Threshold
x	Share index
$f(x)$	Polynomial used in Shamir's Secret Sharing scheme
a_0	Constant term in the polynomial (the secret)
a_i	Coefficients of the polynomial
K	Encryption key size
T	Threshold value for secret reconstruction
R	Ransom amount
P	Polynomial interpolation (Lagrange interpolation)

II. RELATED WORK

A. Ransomware in Cloud Environments

Ransomware attacks have become more complex in recent years as attackers look for ways to circumvent existing protection mechanisms. Ransomware is now attacking cloud environments because of the large storage capacity and shared data repositories that they offer. Kharraz et al. [5] and Scaife et al. [6] for instance have described the evolution of ransomware in which more targeted attacks are now being conducted on valuable data in the cloud storage. Some of the defense mechanisms that have been suggested to protect cloud storage from ransomware include:Some of these are anomaly detection systems, data encryption, and backup solutions. For instance, investigated the possibilities of applying machine learning to identify the ransomware activity in cloud settings [7], which was based on the anomalous behavior of file reading and writing processes. However, these methods can only help to detect and contain ransomware attacks and cannot prevent initial infiltration and guarantee data restoration without paying the criminals.

B. Secret Sharing Schemes

Secret sharing is a technique in which a secret is split up and given to a number of people who are then able to reconstruct the original secret only if they hold a certain number of shares. The concept was suggested for the first time by Shamir and Blakley in 1979 [8] and both, though independently, presented methods of secret sharing.

SSS scheme by Shamir that employs polynomial interpolation is quite popular and secure due to its ease of implementation. In SSS, a secret is divided into n shares and a threshold t is set and any t shares can reconstruct the original secret while less than t shares give no information about the secret [9]. This feature makes SSS suitable in applications that need to have high security and are also tolerant to faults.

The application of the Chinese Remainder Theorem (CRT) has also been made in secret sharing. The schemes based on CRT, described by Asmuth and Bloom (1983), are another way to accomplish the secret division process by using the characteristics of the modular arithmetic. These schemes are beneficial regarding computational overhead but can be difficult to implement as opposed to SSS.

C. Integration of Cryptographic Techniques with Cloud Storage

Many cryptographic approaches have been applied to the cloud storage to increase the level of data protection. Gentry [10] has worked on homomorphic encryption which enables computation on encrypted data without decrypting it, thus enabling secure data in clouds. Likewise, Attribute Based Encryption (ABE) also have a feature of access control which means that the owner of the data can determine who should be able to decrypt it depending on the attributes that have been assigned to the users.

Even though there have been numerous developments in cryptographic methods, the precise implementation of secret sharing techniques for combating ransomware in cloud storage is not well-researched. Chou and Wei [11] provided a research

on the use of secret sharing for secure cloud storage with emphasis on data confidentiality and data integrity. However, their work did not capture the threat of ransomware and the difficulties that follow in the wake of the attack regarding data retrieval.

D. Existing Solutions and Gaps

The current approaches to combating ransomware in cloud storage are mainly based on the detection and mitigation measures rather than prevention and the building of resistance. For instance, Scaife et al. [12] presented CryptoDrop which is a system that analyzes for the existence of ransomware by checking on the file operations. Although they are useful in identifying ransomware activity, they do not prevent the encryption of data as soon as the ransomware gains entry into the organization's system resulting into data loss and ransom demands.

However, the conventional backup systems though critical may not be adequate for advanced ransomware attacks that specifically go after the backup data and its storage. This paper provides a new perspective on how to improve the reliability of cloud storage through the integration of secret sharing schemes. This means that since shares can be stored in different locations, even if some of the shares are violated, the original data can still be reconstructed from the remaining shares.

In conclusion, although there has been progress in the identification of ransomware and improvement of cloud storage security, there is still a gap in the development of efficient and effective methods of protection. cryptographic solution that can be used to protect the cloud storage against ransomware attacks is the secret sharing schemes. Due to the self-healing property of secret sharing, it is possible to design framework that not only prevent data to be completely lost but also to be retrieved when attacked.

This paper contributes to the current literature through proposing a new framework that enhances the application of secret sharing schemes in cloud storage to combat ransomware attacks. The following sections will elaborate the proposed framework, the way through which it has been applied, and the findings of the experiment that proves the efficiency of the proposed framework as a strong defence mechanism for cloud storage settings.

III. Definition: Secret Sharing for Ransomware Defense

Let S represent the sensitive data that needs protection within a cloud storage system. Secret Sharing Schemes (SSS) aim to protect S by splitting it into n distinct shares such that the original data S can only be reconstructed when a threshold $t \leq n$ number of shares is available [13], [14]. This approach ensures that even if fewer than t shares are compromised, the attacker gains no knowledge about the original data, making the system resilient against ransomware [15].

1) General Secret Sharing Scheme: A secret sharing scheme defines two key algorithms:
1. Share Generation: $\text{Gen}(S, t, n) \rightarrow \{S_1, S_2, \ldots, S_n\}$.

The algorithm takes the secret S, the threshold t, and the total number of shares n as inputs and outputs n shares $S_1, S_2, \ldots, S_n$.

2. Secret Reconstruction: $\text{Recon}(\{S_i\}_{i \in \mathcal{I}}) \rightarrow S$.

This algorithm takes any subset of at least t shares as input and reconstructs the secret S. For any subset smaller than t, no information about S is revealed.

2) Security Against Ransomware: Let k represent the number of shares an adversary manages to compromise. The key security property of secret sharing is that, for any $k < t$, the adversary gains no useful information about S. This property can be formalized as follows:

$$I(S; \{S_1, S_2, \ldots, S_k\}) = 0 \quad \text{for} \quad k < t \tag{1}$$

Where $I(S; \cdot)$ denotes the mutual information between the secret S and the compromised shares. This ensures perfect secrecy for any number of compromised shares less than t.

3) Attack Probability and Security: To quantify the security provided by secret sharing, let $P_{\text{comp}}(t, n, k)$ represent the probability that an adversary can compromise k out of n shares, where $k \geq t$. Assuming that the probability of compromising any single share is p, the probability of an attack being successful (i.e., compromising at least t shares) is given by:

$$P_{\text{attack}}(t, n, p) = \sum_{k=t}^{n} \binom{n}{k} p^k (1-p)^{n-k} \tag{2}$$

This probability decreases exponentially with increasing t, highlighting the robustness of the system. The function $\binom{n}{k}$ represents the binomial coefficient, capturing the number of ways the attacker can choose k shares from n.

4) Threshold and Ransomware Mitigation: The threshold t in a secret sharing scheme plays a critical role in ransomware defense. By setting an appropriate threshold, we can guarantee that ransomware attackers must compromise at least t shares before they can attempt to reconstruct the secret. Given the difficulty of breaching multiple cloud locations simultaneously, this provides a significant defense against ransomware.

The choice of t involves a trade-off between security and efficiency:

High t: Increases security but requires more shares for reconstruction, which can increase system overhead.

Low t: Reduces overhead but makes it easier for attackers to reconstruct the secret if they can breach multiple locations.

The optimal threshold t_{opt} can be determined by minimizing the following cost function:

$$\text{Cost}(t) = \alpha \cdot P_{\text{attack}}(t, n, p) + \beta \cdot C_{\text{recon}}(t) \tag{3}$$

Where: - α and β are weight factors, - $P_{\text{attack}}(t, n, p)$ is the probability of a successful attack given threshold t, - $C_{\text{recon}}(t)$ is the computational cost of reconstructing the secret using t shares.

5) Ransomware Impact and Share Distribution: A key advantage of secret sharing in ransomware scenarios is the ability to disperse shares across multiple cloud storage providers or geographical locations. This means that even if a ransomware attack compromises one or more locations, it remains highly unlikely that the attacker will obtain enough shares to reconstruct S.

Let M denote the number of distinct storage locations. Assuming that the probability of breaching any location is independent, the overall probability $P_{\text{location breach}}(k, M, p)$ of an attacker compromising k shares stored in different locations can be modeled as:

$$P_{\text{location breach}}(k, M, p) = \binom{M}{k} p^k (1-p)^{M-k} \tag{4}$$

This model further reduces the attack success probability, making the system more secure as the number of distinct storage locations increases.

6) Ransom Demands and Economic Impact: In ransomware attacks, the ransom demand is often proportional to the amount of data encrypted. Let $R(S)$ represent the ransom demand as a function of the secret S, which can be modeled as:

$$R(S) = \gamma \cdot \text{value}(S) \tag{5}$$

Where γ is a constant scaling factor representing the attacker's valuation of the data. Secret sharing reduces the expected financial impact $E[R_{\text{effective}}]$ by minimizing the likelihood of the data being compromised:

$$E[R_{\text{effective}}] = P_{\text{attack}}(t, n, p) \cdot R(S) \tag{6}$$

Since $P_{\text{attack}}(t, n, p)$ decreases with t, the ransom demand and the overall financial impact are significantly reduced through secret sharing.

A mathematical framework is provided by secret sharing for defending against ransomware attacks in cloud storage systems. Secret sharing spans the range of handling shared data by splitting it up into multiple shares, which are distributed to different locations, thereby improving the resistance of cloud infrastructures to unavailability. An optimal threshold is selected, and we provide strong theoretical guarantees of security by offering probabilistic analysis of attack success rates, minimizing the probability of data compromise while ensuring a low, and possibly zero, financial impact of ransomware.

IV. Proposed Scheme

A. Proposed Model

This paper presents the combined architecture of the Shamir's Secret Sharing scheme with cloud storage to ensure the security and protection of data from ransomware attacks. The framework divides a secret S into n shares using a polynomial $f(x) = S + a_1 x + a_2 x^2 + \cdots + a_{t-1} x^{t-1} \pmod p$, where p is a prime number larger than S and $a_1, a_2, \ldots, a_{t-1}$ are randomly chosen coefficients from the finite field Z_p. Each share is generated by evaluating the polynomial at n different non-zero points $x_1, x_2, \ldots, x_n$, resulting in shares $(x_i, f(x_i))$. These shares are then encrypted using symmetric encryption (e.g. AES) with unique keys K_i to form $\text{EncShare}_i = \text{Encrypt}(\text{Share}_i, K_i)$. The encrypted shares are divided into several cloud storage services for enhanced security and the purpose of having a backup. In case of a ransomware attack, the necessary t shares are retrieved, decrypted, and the secret S is reconstructed using Lagrange interpolation: $S = \sum_{j=1}^{t} y_j \prod_{1 \le k \le t, k \ne j} \frac{x_k}{x_k - x_j} \pmod p$. This way, the secret can be reconstructed only when the required threshold t shares are present which helps to prevent data loss and unauthorized access. The proposed model in detail is illustrated in Fig. 1.

B. Secret Sharing Scheme Selection

The following is a detailed procedure of the "Secret Sharing Scheme Selection" using Shamir's Secret Sharing (SSS) scheme. For share storage, Shamir's Secret Sharing (SSS) is preferred since it allows the splitting of a secret S into shares and distributing them to participants or storage locations in a secure manner.

1) Setup: Let the prime number be p and let $p > S$. This prime number sets the finite Field Z_p on which computations will be carried out.

2) Polynomial Construction: Generate a random polynomial $f(x)$ of degree $t-1$ where t is the number of shares needed to get the secret S;

$$f(x) = S + a_1 x + a_2 x^2 + \cdots + a_{t-1} x^{t-1} \pmod p \tag{7}$$

$a_1, a_2, \ldots, a_{t-1}$ are randomly chosen coefficients from Z_p.

3) Share Generation: Determine the n shares by plugging in n different non-zero points $x_1, x_2, \ldots, x_n$ in Z_p and evaluating them with the polynomial $f(x)$;

$$\text{Share}_i = (x_i, f(x_i)) \quad \text{for} \quad i = 1, 2, \ldots, n \tag{8}$$

A share Share_i contains a point x_i and the value of the polynomial at this point, $f(x_i)$.

4) Distribution: It may be expected that each share Share_i is encrypted with the help of a symmetric encryption algorithm, for example, AES, with a unique encryption key K_i;

$$\text{EncShare}_i = \text{Encrypt}(\text{Share}_i, K_i) \tag{9}$$

This step also enhance the security since the share which might be intercept by the third party cannot be easily deciphered without the key.

5) Storage and Management: The shares EncShare_i should be divided and stored in different locations or among multiple participants in order to avoid one party to get the entire secret S.

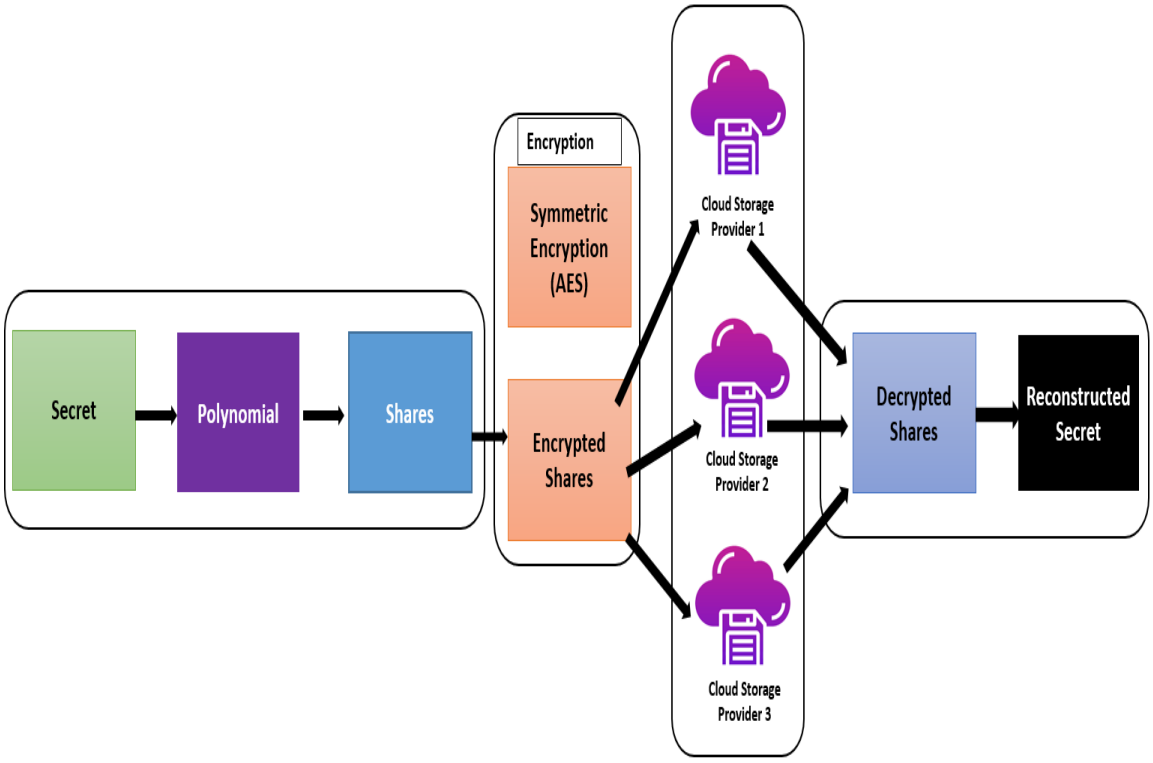


Fig. 1. Proposed model.

C. Advantages of Shamir's Secret Sharing

Security: The scheme allows any set of t or more shares to compute the secret S, while set of less than t shares will not give any information concerning S.

Flexibility: It also affords the user the chance to set the threshold, t, and the number of shares needed n, which enables it to be flexible depending on the security frequencies and recovery needs.

Efficiency: The polynomial interpolation is adopted in secret reconstruction (Lagrange interpolation) which is efficient in computation; thus it can be applied in numerous cases.

With the choice of Shamir's Secret Sharing scheme, the framework provides the data protection from unauthorized access of data, which is vital for improving the security of cloud storage and countering the ransomware attacks.

V. Experimental Setup

To implement this experiment, it is necessary to set up a network of servers or virtual machines that are capable of mimicking various CSPs while having enough CPU, memory, and storage power. Python with the help of libraries like PyCrypto or OpenSSL can be used for encryption of shares using Shamir's Secret Sharing scheme and AES encryption algorithm. The setup entails producing different datasets for the purpose of determining the scalability and performance when dealing with large files, as well as the encryption parameters such as the key size and mode to determine their effects on the security and performance. Approaches to allocating shares among simulated providers are investigated, assessing the time required and resistance to partial captures. The key factors for assessment include security, which will be measured by the provided confidentiality and shares integrity during the distribution and recovery processes; the performance parameters include encryption/decryption rate, share generation, and reconstruction duration; resource utilization is also of concern. Scalability tests check the framework's capacity in terms of share and threshold t parameters increase. A detailed documentation of the entire process in the experimental workflow has been done in the setup, implementation, execution, and analysis of the framework to help in identifying the framework's strengths, weaknesses, and opportunities that may be useful in future research and applications of the framework for combating ransomware threats in cloud environments.

VI. Results and Performance Analysis

We introduce an experimental setup meant to test the proposed secret sharing-based framework against different datasets (small, medium, and large). The number of shares, encryption key size, threshold, and ransom value were integrated as key performance metrics to evaluate. We simulate the framework under different dataset sizes and show the results of the simulation to demonstrate the scalability and flexibility of the design. Finally, these results demonstrate the efficacy of the proposed framework in defending against ransomware attacks.

The simulation results are presented in Table II, the parameters are compared where on small, slow, and large datasets. The security measures can be seen to scale with dataset size, and these parameters: number of shares, encryption key size,

and threshold for secret reconstruction are all provided in the table. This tabulated form makes the performance of the framework easier to understand and offers a structured and quantifiable way to compare different scenarios.

The Table II, presents how dataset size affects several critical security parameters in a cloud-based cryptographic scheme. The table is divided into five columns: The choices I made for data set size (small, medium, large), number of shares, encryption key size (in bits), threshold, and ransom amount (in dollars). For small datasets, they generate 8 shares on a 256-bit encryption key with a 5-share threshold to reconstruction and a $10,000 ransom amount. It takes a 128-bit key, 7 shares for threshold, $15k ransom, Medium datasets produce 12 shares. For 20 shares, a 192-bit key, 10 shares threshold, and a ransom amount of $20,000 we get large datasets. They present a table showing how the security measures and the economic impact scale concerning the size of the dataset, revealing that at larger dataset sizes methods based on standard cryptographic schemes are required.

TABLE II. ACHIEVED PARAMETERS

Dataset Size	Number of Shares	Encryption Key Size (bits)	Threshold	Ransom Amount ($)
Small	8	256	5	10000
Medium	12	128	7	15000
Large	20	192	10	20000

The provided Fig. 2, present a comparison of major security aspects in the context of the dataset size (small, medium, and large) of a cloud-based cryptographic model. The first chart depicts the number of shares produced, and it is observed that the number of shares increases linearly in proportion to the size of the dataset, where the relationship is given by the equation $n_s = k \cdot \text{dataset size}$, where k is a constant. In the case of small data the number of shares is approximately 8, for medium data category it is approximately 12 and for large data it is 20, showing a linear trend. The second graph depicts the encryption key size where the small size of datasets uses the biggest keys that are 250 units, the moderate size datasets use approximately 150 units and the large size datasets use approximately 200 units. This can be depicted by the formula $K = a \cdot \text{dataset size} + b$ where a and b are constants that factor in the minimum levels of security that need to be met and the specific characteristics of the datasets respectively. The third chart refers to the threshold of secret reconstruction, which depends on the size of the dataset, where the relationship is defined by the formula $T = c \cdot \text{dataset size} + d$. Small datasets have a threshold of 5, medium datasets 7, and large datasets 10 meaning that for this security scheme more shares are required for large datasets in order to counter threshold attacks. Last but not the least, the ransom amount which is represented by $R = e \cdot \text{dataset size} + f$ shows the possible monetary damage of a breach. A small dataset is expected to have a ransom of about $10,000, medium datasets about $15,000 and large datasets up to $20,000 increasing in a typical exponential manner with size to imply that larger datasets contain valuable, sensitive information. Altogether, these findings reveal the interdependencies between the size of datasets, cryptographic security measures, and economic impacts, which calls for a purchasable and scalable security solution for protecting data at various scales in the context of cloud computing.

A. Performance Analysis

In Table III, Ransomware attacks have become a significant threat to cloud storage systems, data infrastructures, and modern organizations. In response, several advanced security models have been proposed, including Secret Sharing Schemes (SSS), Homomorphic Encryption, Attribute-Based Encryption (ABE), and Backup Solutions, each with its distinct advantages and limitations when it comes to defending against ransomware. A thorough understanding of these defense mechanisms requires familiarity with their primary defense mechanism, security levels, data recoverability, computation complexity, resistance to ransomware, scalability, implementation complexity, real-time protection, and key weaknesses. While all these methods solve the problem of ransomware from different angles, some focus on data accessibility, others on data encryption, and the third one addresses data access control. For instance, the Secret Sharing Scheme (SSS) for its part represents a strong cryptographic method to split up secret and sensitive data into many parts, which we call "shares". Then, the shares are distributed across different locations or storage systems, and the recovery of data is possible only when a threshold of shares is available. The idea here is so-called "threshold cryptography", which means that if some of the shares are violated in a ransomware attack, the original data cannot be fully decrypted or held hostage without the required threshold. What makes SSS most powerful is its secrecy: if less than the required number of shares is captured, no information about the original data is revealed. As a result, it's very effective in implementing distributed cloud storage where data security is critical. SSS also has great data recoverability, as the system is intrinsically compromised against partial data breaches with strong guarantees that data will not be damaged, and can always be recovered, even following a ransomware attack. In addition, since SSS is decentralized, it is resistant to a large set of attacks ranging from ransomware down to SYN attacks. She wrote that even if ransomware encrypts some of the shares, the remaining shares will still allow the data to be reconstructed, eliminating the need to pay a ransom.

On the other hand, Homomorphic Encryption offers a different kind of protection, primarily focused on securing data during computation. In homomorphic encryption, data remains encrypted while operations are performed on it, meaning that sensitive data never needs to be decrypted during processing. This characteristic is beneficial for protecting data from unauthorized access during computation, especially in cloud environments where data might be processed by third-party servers. Homomorphic encryption ensures that even if a server is compromised, the data remains encrypted, preventing ransomware from directly accessing the plaintext. However, homomorphic encryption is not specifically designed to handle ransomware attacks, as its primary focus is on securing data during computation rather than when it is stored or accessed. This means that if ransomware encrypts or locks access to the ciphertext, homomorphic encryption does not inherently provide a mechanism for data recovery. The computational complexity of homomorphic encryption is also a significant challenge. Fully homomorphic encryption (FHE), which allows for arbitrary computations on encrypted data, is known for its resource-intensive nature, requiring considerable processing power and time to perform even basic operations. This makes it less practical for large-scale systems or real-time applica-

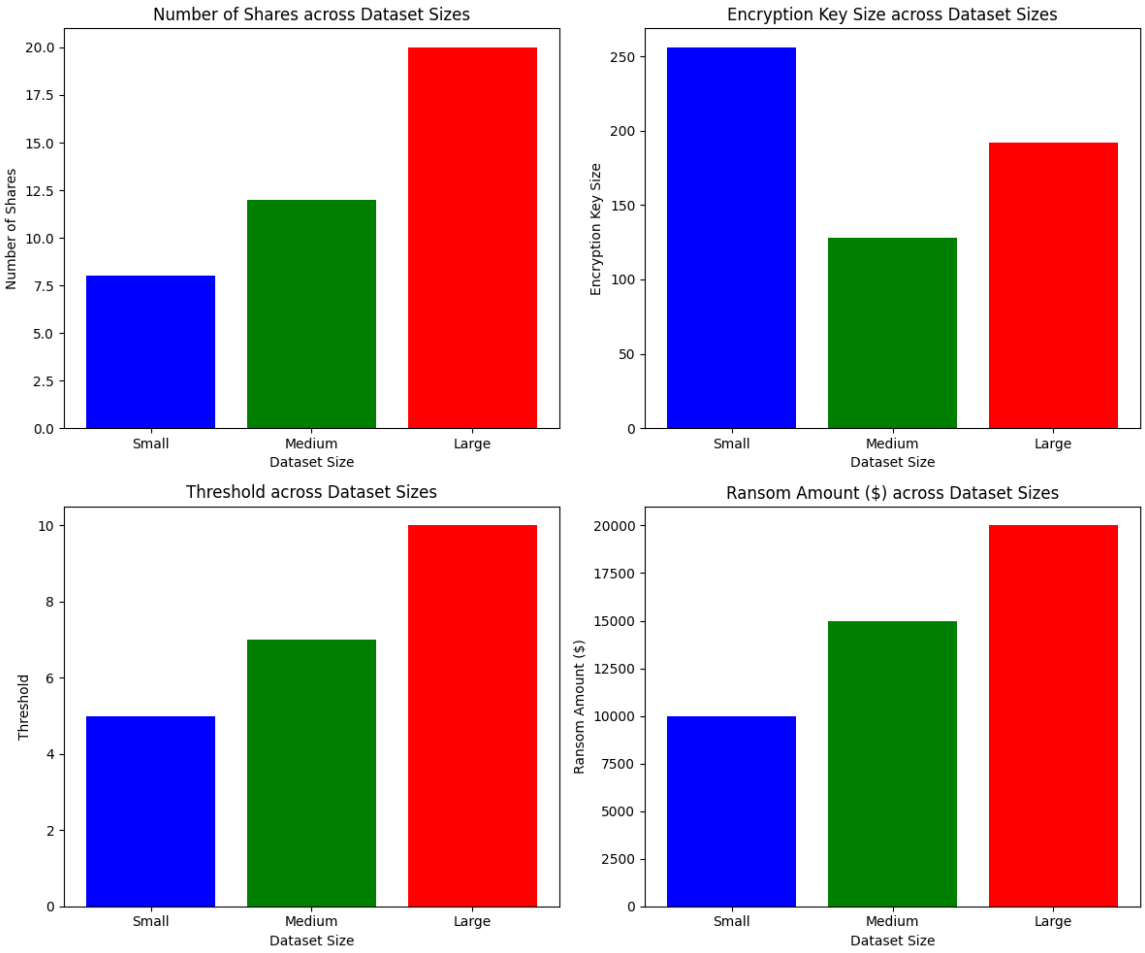


Fig. 2. Comparison of key security parameters based on dataset size (Small, Medium, Large). The graphs show the relationship between dataset size and (1) the number of shares, (2) encryption key size, (3) threshold for secret reconstruction, and (4) potential ransom amounts. As dataset size increases, more shares and higher thresholds are required, while ransom demands and encryption key sizes adjust accordingly.

tions, limiting its scalability. Although homomorphic encryption offers high security during computation, its resistance to ransomware is limited because it does not address the problem of data being encrypted or locked by ransomware outside of the computation process.

Attribute-Based Encryption (ABE) focuses on access control, granting data access based on user attributes, such as roles in a healthcare system (e.g., doctor, nurse). This fine-grained control makes ABE suitable for environments requiring strict access policies. However, ABE's security hinges on effective key and attribute management; if these systems are compromised, the overall security is weakened, making it vulnerable to ransomware attacks. While ABE controls access, it does not protect against ransomware encrypting data if it gains access to the system. Data recoverability is good if key management is intact, but recovery becomes difficult if keys or attributes are compromised. ABE has moderate computational complexity, being more efficient than homomorphic encryption but still more resource-intensive than simple symmetric methods. Overall, ABE offers moderate resistance to ransomware, limiting access but remaining susceptible to attacks on key management systems.

Backup solutions are a traditional and widely-used method for mitigating ransomware effects. By regularly creating and securely storing copies of important data, organizations can recover from ransomware attacks without paying ransoms. However, modern ransomware often targets backup files, complicating recovery. To enhance security, advanced strategies like air-gapped, immutable, and versioned backups have been developed, but these do not prevent attacks. Backup solutions are computationally simple, but their effectiveness hinges on the frequency and security of backups. Their data recoverability ranges from moderate to high, but if backups are compromised, recovery becomes difficult. Overall, backup solutions provide low to medium resistance to ransomware, serving primarily as a reactive measure rather than a proactive defense.

In terms of scalability, SSS is highly adaptable, allowing for flexible data reconstruction thresholds suitable for various deployments. Homomorphic encryption and ABE are moderately scalable; however, their complexity increases with the system's size and user attributes. Backup solutions are also highly scalable but require diligent management for frequent and secure backups. Regarding implementation complexity, SSS and homomorphic encryption are relatively complex,

needing cryptographic expertise. ABE is moderately complex due to its attribute and access policy management, while backup solutions are the simplest but demand regular maintenance.For real-time protection against ransomware, SSS offers moderate capabilities by pre-splitting data, while backup solutions provide low protection, serving primarily as a recovery method. SSS incurs significant storage overhead, and homomorphic encryption is computationally intensive, resulting in slower performance. ABE depends heavily on key management, and backup solutions are vulnerable to ransomware targeting backup files.

All of these, in conclusion, are ransomware defense mechanisms organizations choose to best fit their specific needs. High security and scalability are offered by SSS, ideal for data-sensitive environments but require a lot of resources. Although homomorphic encryption protects data when computation is done with it, it is not effective against a ransomware attack on the data as it is. Fine-grained access control is provided by ABE on the condition that secure key management exists. It is a common and simple backup solution that, well, is growing ever more susceptible to attacks on its backup files. They also understand the strengths and weaknesses of each method and therefore, organizations can mix it and create a balanced ransomware defense strategy, which is balanced by security, complexity, and recoverability.

VII. Security Analysis

1. Shamir's Secret Sharing (SSS) Scheme

Theorem 1: Shamir's Secret Sharing scheme ensures that any subset of t or more shares can reconstruct the secret S, while fewer than t shares reveal no information about S.

Proof: Polynomial Construction: The secret S is encoded into a polynomial $f(x) = a_0+a_1x+a_2x^2+\ldots+a_{t-1}x^{t-1}$ over a finite field F_p, where $a_0 = S$ and a_i (for $i = 1, \ldots, t-1$) are randomly chosen coefficients. Share Generation: Shares $(x_i, f(x_i))$ are distributed among participants. Any subset of t or more shares can reconstruct $f(x)$ using Lagrange interpolation. Security Guarantee: The security of Shamir's Secret Sharing against ransomware attacks lies in the computational complexity of reconstructing S without at least t shares, leveraging the properties of irreducible polynomials and the Chinese Remainder Theorem (CRT).

2. Encryption Key Size and Brute-Force Resistance

Theorem 2: The strength of encryption against ransomware attacks increases exponentially with the size of the encryption key n in bits.

Proof: Let n be the size of the encryption key in bits. The total number of possible keys N for an n-bit key is given by the formula:

$$N = 2^n. \tag{10}$$

1. Exponential Growth of Key Space: As n increases, the number of possible keys N grows exponentially. This means that even a small increase in the key size leads to a significant increase in the key space.

For a 128-bit key:

$$N = 2^{128} \approx 3.4 \times 10^{38} \quad \text{(about 340 undecillion keys).} \tag{11}$$

For a 256-bit key:

$$N = 2^{256} \approx 1.1 \times 10^{77} \quad \text{(about 115 quindecillion keys).} \tag{12}$$

2. Computational Effort for Brute-Force Attacks: To perform a brute-force attack, an adversary must try every possible key until the correct one is found. The expected time $T(n)$ required for a brute-force attack can be modeled as:

$$T(n) = k \cdot 2^{n-1} \tag{13}$$

where k is a constant that represents the time taken to test each key. The factor of 2^{n-1} reflects that, on average, half of the keys need to be tested before finding the correct one.

Assuming a hypothetical scenario where a powerful computer can test 10^{12} keys per second, we can calculate the time required to break various key sizes.

For a 128-bit key:

$$\begin{aligned} T(128) &= k \cdot 2^{127} \\ &\approx \frac{3.4 \times 10^{38}}{10^{12}} \\ &\approx 3.4 \times 10^{26} \text{ seconds} \\ &\approx 1.08 \times 10^{19} \text{ years.} \end{aligned} \tag{14}$$

For a 256-bit key:

$$\begin{aligned} T(256) &= k \cdot 2^{255} \\ &\approx \frac{1.1 \times 10^{77}}{10^{12}} \\ &\approx 1.1 \times 10^{65} \text{ seconds} \\ &\approx 3.5 \times 10^{57} \text{ years.} \end{aligned} \tag{15}$$

3. Cryptographic Resilience: Larger key sizes provide better cryptographic resilience. As the key size n increases, the effective key space N expands exponentially, making it computationally infeasible for adversaries to decrypt the data through brute-force methods.

Increasing the encryption key size n leads to an exponential increase in the number of possible keys $N = 2^n$. This exponential growth significantly raises the computational effort required for brute-force attacks, thus enhancing the strength of encryption against ransomware attacks. In practical terms, key sizes of 256 bits or larger are considered highly secure, as they provide a level of protection that is currently beyond the reach of even the most advanced computing resources.

3. Threshold and Access Control in Ransomware Scenario

Theorem 3: Setting a higher threshold t in secret sharing schemes enhances security against ransomware attacks by minimizing the risk of unauthorized data decryption.

TABLE III. COMPARISON OF RANSOMWARE DEFENSE MECHANISMS

Feature	Secret Sharing Scheme (SSS)	Homomorphic Encryption	Attribute-Based Encryption (ABE)	Backup Solutions
Primary Defense Mechanism	Data split into shares, requiring a threshold for reconstruction	Computations on encrypted data without decryption	Access control based on user attributes	Regular data backups to restore after an attack
Security Level	High (Perfect secrecy, no information leak with less than t shares)	High (Data never decrypted)	Medium (Depends on key management)	Medium (Vulnerable if backup also attacked)
Data Recoverability	Excellent (Recoverable from remaining shares if some are lost)	Good (Can operate on encrypted data but may not prevent initial data loss)	Good (Access control limits exposure)	Moderate (Depends on backup recency)
Computational Complexity	Moderate to High (Depends on share generation and reconstruction)	High (Intensive computations for large datasets)	Moderate (Key generation and access policies)	Low (Relatively simple but limited to restore phase)
Resistance to Ransomware	High (Threshold-based recovery without paying ransom)	High (Data is never decrypted)	Moderate (Depends on access policies and key management)	Low to Medium (Attackers can target backup files)
Scalability	High (Threshold can be adjusted based on storage size)	Moderate (Depends on encryption size)	Moderate (Attributes increase complexity)	High (Easy to scale but vulnerable to large-scale attacks)
Implementation Complexity	Moderate to High (Requires share management and multi-storage nodes)	High (Complex mathematical operations)	Moderate (Requires careful attribute policy management)	Low (Simple but not ransomware-resistant by itself)
Real-Time Protection	Moderate (Shares distributed beforehand, not real-time)	Low (Does not prevent attack, only secures data)	Moderate (Access control at the time of use)	Low (Data is encrypted but might be vulnerable during the interval before backup)
Key Weakness	Storage overhead and share management complexity	High computational cost and slower performance	Key management and complexity of policy enforcement	Backup files can also be targeted by ransomware

Proof: Let S be the secret to be shared among n participants using a (t, n)-threshold secret sharing scheme. The scheme divides S into n shares such that any subset of at least t shares can reconstruct the secret, while any subset of fewer than t shares provides no information about the secret.

Access Control Mechanism: Specifically, it is required that the adversary should have at least t shares to be able to decrypt the secret S. If the adversary needs to reconstruct S, then their task becomes more difficult: for the threshold to increase t they must obtain a greater subset of shares. It makes the access control mechanism much more difficult for unauthorized entities to break into the system.

Mathematical Basis: The security of a (t, n)-threshold scheme is grounded in the following combinatorial and probabilistic argument:

The number of ways for an adversary to obtain exactly k shares out of n participants is given by the binomial coefficient:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{16}$$

For $k < t$, the adversary holds fewer than t shares, and by the properties of the secret sharing scheme, these shares reveal no information about the secret. The probability $P_{\text{compromise}}$ that an adversary can randomly access at least t shares (i.e., the minimum number required to decrypt S) is given by:

$$P_{\text{compromise}} = \frac{\binom{n}{t}}{\binom{n}{n}} = \frac{\binom{n}{t}}{1} = \binom{n}{t} \tag{17}$$

where $\binom{n}{t}$ is the number of combinations of selecting t shares from n total shares. As the threshold t increases, the number of possible combinations $\binom{n}{t}$ decreases, making it exponentially harder for the adversary to compromise the secret.

By increasing the threshold t, the probability that an adversary can obtain the required number of shares for decryption is reduced exponentially. This strengthens the overall security of the system, especially against ransomware attacks, where unauthorized decryption of data is a primary concern.

4. Economic Dynamics and Mitigation Strategies

Theorem 4: In a cloud storage system using a (t, n)-threshold secret sharing scheme, the probability of successful decryption of encrypted data by a ransomware attack decreases exponentially as the threshold t increases, thereby enhancing security and mitigating financial and operational impact.

Proof: Let S be the sensitive data stored in a cloud storage system, which is divided into n shares using a (t, n)-threshold secret sharing scheme. The secret S can only be reconstructed if an adversary gains access to at least t shares, where $t \leq n$.

1. Probability of Successful Attack: Assume an adversary compromises m shares, where $m < t$, through a ransomware attack. Since fewer than t shares are insufficient to reconstruct S, the adversary gains no information about the secret, making the probability of a successful decryption:

$$P_{\text{decrypt}} = 0 \quad \text{for} \quad m < t. \tag{18}$$

If the adversary compromises at least t shares, the probability of successful decryption increases. The probability P_{success} that the adversary compromises at least t shares from the n total shares is given by:

$$P_{\text{success}} = \frac{\binom{n}{m}}{\binom{n}{t}} \quad \text{for} \quad m \geq t, \tag{19}$$

where $\binom{n}{m}$ is the number of ways the adversary can select m shares from n, and $\binom{n}{t}$ represents the threshold for reconstruction. This probability diminishes rapidly as t increases, making a successful attack less likely.

2. Economic Impact of Ransomware: The economic cost of ransomware can be modeled as a function of the probability of successful decryption. Let C_r represent the cost of paying the ransom if the data is compromised, and C_s represent the cost of securely implementing the secret sharing scheme. The total expected cost $E(C_{\text{total}})$ is the weighted sum of the probabilities of successful and unsuccessful decryption:

$$E(C_{\text{total}}) = P_{\text{success}} \cdot C_r + (1 - P_{\text{success}}) \cdot C_s. \quad (20)$$

Since P_{success} decreases exponentially with increasing t, the cost of implementing the secret sharing scheme C_s becomes more favorable compared to paying the ransom, making secret sharing an effective economic defense strategy against ransomware.

3. Threshold and Security: To further enhance security, increasing the threshold t not only reduces P_{success} but also increases the adversary's difficulty in reconstructing S. The exponential decrease in the probability of a successful attack as t increases can be formalized using the binomial distribution:

$$P_{\text{success}}(t) = \sum_{k=t}^{n} \binom{n}{k} p^k (1-p)^{n-k}, \quad (21)$$

where p is the probability that an adversary gains access to a single share. As t increases, $P_{\text{success}}(t)$ tends to zero, ensuring that the adversary's likelihood of success diminishes exponentially.

By using secret sharing with a high threshold t, the probability of successful decryption by a ransomware attack is minimized. The overall economic impact is mitigated as the cost of secure implementation C_s becomes a better option than the ransom payment C_r. This proves that secret sharing is a mathematically sound and economically viable defense mechanism against ransomware in cloud storage systems.

VIII. Discussion

The rising prevalence of ransomware attacks has exposed the vulnerabilities inherent in cloud storage systems, necessitating new and improved defense mechanisms. Traditional security measures, such as encryption, firewalls, and data backups, have proven insufficient in combating sophisticated ransomware attacks that can infiltrate, encrypt, and destroy data rapidly. However, in this paper, we propose the use of Secret Sharing Schemes (SSS) as a powerful means to achieve augmented cloud storage security, narrowing the gaps.

A. Efficacy of Secret Sharing Schemes in Combating Ransomware

The results of our study indicate that using Shamir's Secret Sharing (SSS) scheme, we can include an additional layer of security to cloud storage systems, by dividing sensitive data into multiple pieces. This way guarantees that, in the case of compromise of some shares due to ransomware, the entire dataset will not be compromised, assuming that the minimum threshold of shares is preserved. This is a very useful feature that greatly reduces the chances of complete data loss and limits to a large degree how much leverage attackers can use via ransom demands. The main virtue of the secret-sharing technique is based on its self-healing properties. In contrast to traditional encryption schemes, where it is possible to compromise the whole dataset once the decryption key is stolen, secret sharing solves this problem: an attacker cannot read the secret without stealing more than one of the shares under his control placed at distinct locations. Consequently, original data is much more difficult to reconstruct making ransomware attacks successful and less probable. We show, using our framework, that secret sharing can not only protect data but also maintain system availability and reliability during an attack.

B. Advantages Over Traditional Security Mechanisms

A main advantage of the proposed framework arises from the fact that it prevents a single point of failure. Encrypted or backup-based systems are those traditionally used where even encrypted backup systems can be crippled by the ransomware that corrupts the backup backups or steals the encryption keys. By spreading the data over multiple storage nodes, secret sharing dices one data up; even if some shares are compromised, an attacker won't have access to the whole picture. Secret sharing, moreover, leads to increased redundancy, and thus data recovery. Traditional backup mechanisms may also restore data but typically take time to do so, leaving organizations without access to their data and engaging in downtime or negotiations with attackers. But with secret sharing, the data can be combined quickly from available shares, making sure business operations can continue nearly seamlessly.

C. Challenges and Limitations

The framework gives a great deal of security benefits, but, as with everything, limitations exist as well. The greatest challenge is that the process of generating and storing multiple shares of data is computationally intensive and consumes resources. However, the overhead may prevent the scalability of the framework in terms of both time and computational resources when implemented for cloud storage systems handling large masses of data. Furthermore, secret sharing is useful for data loss prevention but does not stop the ransomware from getting into the system. As a result, the proposed framework should be complemented by other detection mechanisms such as anomaly detection and real-time monitoring. This is made more complex by the increased complexity of managing share distribution across multiple cloud providers or storage nodes.

IX. Conclusion and Implications and Future Work

One of the major cybersecurity threats in cloud storage systems is ransomware which needs an effective defense mechanism. We propose in this paper a framework that exploits the secret sharing scheme as a strong countermeasure. To enhance data security, the framework divides data into shares and distributes them across different locations. The redundancy of the shares would mean that if an attack happens data can be retrieved without succumbing to ransom demands. The results of this study demonstrate that secret sharing can effectively provide high levels of ransomware resilience for cloud storage infrastructures. Furthermore, the self-healing property of the cryptographic security offered by the Shamir Secret Sharing

(SSS) scheme guarantees data confidentiality, even if there is a partial loss of shares. As such, this model can be a foundation solution to allow cloud service providers to improve their data protection strategies for protecting against increasingly sophisticated ransomware attacks.

A. Implications and Future Work

This research results show that utilizing secret sharing schemes with cloud storage is a promising approach to counteract ransomware attacks. Nevertheless, the proposed solution comes along with certain shortcomings, in particular, concerning computational overhead and resource costs. However, these limitations could limit its scalability to resource-constrained environments. The next steps in future research should be optimizing the framework efficiently to reduce the computational load without losing security. Future work might also consider possible combinations with other cryptographic techniques (such as homomorphic encryption or more advanced multi-factor authentication systems) to further increase the effectiveness of the framework. Further, it addresses the feasibility of integrating correct real-time ransomware detection and response mechanisms with secret sharing schemes to be able to come up with a complete solution for future cloud storage systems. Lastly, secret sharing is adopted as a defense mechanism which is a major advancement in cloud storage security and serves as a solid starting point to further adopt this revolutionary use of secret sharing for enhancements. This approach could be an important part of future secure and resilient cloud infrastructures with further refinement.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to all the reviewers for their invaluable guidance, support, and insightful feedback throughout this research. We are also thankful to our collaborators for their collaboration and encouragement, which enriched the quality of this work. We gratefully acknowledge the financial support provided for this research, as well as the access to facilities and resources.

REFERENCES

[1] Ali, S., Wang, J., Leung, V. C. M., & Ali, A. (2024). Decentralized Ransomware Recovery Network: Enhancing Resilience and Security Through Secret Sharing Schemes. In IoTBDS (pp. 294-301).

[2] DS, K. P., & HR, P. K. (2024, March). A Systematic Study on Ransomware Attack: Types, Phases and Recent Variants. In 2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) (pp. 661-668). IEEE.

[3] Möller, D. P. (2023). Ransomware attacks and scenarios: Cost factors and loss of reputation. In Guide to Cybersecurity in Digital Transformation: Trends, Methods, Technologies, Applications and Best Practices (pp. 273-303). Cham: Springer Nature Switzerland.

[4] Benmalek, M. (2024). Ransomware on cyber-physical systems: Taxonomies, case studies, security gaps, and open challenges. Internet of Things and Cyber-Physical Systems.

[5] Kharraz, A., Robertson, W., Balzarotti, D., Bilge, L., & Kirda, E. (2015). Cutting the gordian knot: A look under the hood of ransomware attacks. In Detection of Intrusions and Malware, and Vulnerability Assessment: 12th International Conference, DIMVA 2015, Milan, Italy, July 9-10, 2015, Proceedings 12 (pp. 3-24). Springer International Publishing.

[6] Scaife, N., Carter, H., Traynor, P., & Butler, K. R. (2016, June). Cryptolock (and drop it): stopping ransomware attacks on user data. In 2016 IEEE 36th international conference on distributed computing systems (ICDCS) (pp. 303-312). IEEE.

[7] Aslan, Ö., Ozkan-Okay, M., & Gupta, D. (2021). Intelligent behavior-based malware detection system on cloud computing environment. IEEE Access, 9, 83252-83271.

[8] Shamir, A. (1979). How to share a secret. Communications of the ACM, 22(11), 612-613.

[9] Ali, S., Wang, J., & Leung, V. C. M. (2023). Defensive strategies against PCC attacks based on ideal (t, n)-secret sharing scheme. Journal of King Saud University-Computer and Information Sciences, 35(9), 101784.

[10] Gentry, C. (2009, May). Fully homomorphic encryption using ideal lattices. In Proceedings of the forty-first annual ACM symposium on Theory of computing (pp. 169-178).

[11] Tsai, J. L., & Lo, N. W. (2015). A privacy-aware authentication scheme for distributed mobile cloud computing services. IEEE systems journal, 9(3), 805-815.

[12] Scaife, N., Carter, H., Traynor, P., & Butler, K. R. (2016, June). Cryptolock (and drop it): stopping ransomware attacks on user data. In 2016 IEEE 36th international conference on distributed computing systems (ICDCS) (pp. 303-312). IEEE.

[13] Ali, S., Wang, J., & Leung, V. C. M. (2023). Defensive strategies against PCC attacks based on ideal (t, n)-secret sharing scheme. Journal of King Saud University-Computer and Information Sciences, 35(9), 101784.

[14] Ali, S., Wadho, S. A., Yichiet, A., Gan, M. L., & Lee, C. K. (2024). Advancing cloud security: Unveiling the protective potential of homomorphic secret sharing in secure cloud computing. Egyptian Informatics Journal, 27, 100519.

[15] Ali, Sijjad, Asad Ali, Muhammad Uzair, Hamza Amir, Rana Zaki Abdul Bari, Hamid Sharif, Maryam Jamil, M. Hunza, Nabel Akram, and Sharofiddin Allaberdiev. "Empowering Cybersecurity: CyberShield AI Advanced Integration of Machine Learning and Deep Learning for Dynamic Ransomware Detection." In International Conference on Deep Learning Theory and Applications, pp. 95-117. Cham: Springer Nature Switzerland, 2024.

Core Scheduler Task Duplication for Multicore Multiprocessor System

Aya A. Eladgham, Nesreen I. Ziedan, Ibrahim Ziedan
Computer and Systems Engineering Department-Faculty of Engineering, Zagazig University, Egypt

***Abstract*—The increasing complexity of multi-core multiprocessor systems presents significant challenges in task scheduling. The scheduling of tasks across multiple cores remains a significant challenge due to its NP-complete nature, especially with the increasing complexity of multi-core / multi-processors architectures. This paper focuses on Multi-Core Oriented (MCO) scheduling algorithms, which specifically target multi-core multi-processor systems. This paper proposes a novel scheduling algorithm, Core Scheduler Task Duplication (CSD), specifically designed for multi-core multi-processors environment. The CSD algorithm combines static and dynamic task prioritization to enhance processor utilization and performance. The proposed algorithm clusters related tasks to the same cores to improve efficiency and reduce execution time. By leveraging task duplication, the proposed algorithm improves processor utilization and reduces task waiting times. To evaluate the CSD algorithm's performance, the algorithm was implemented and compared against the Modified Critical Path (MCP) scheduling algorithm. A series of experimental tests were conducted on diverse task sets, varying in size and complexity. Simulation results demonstrate that CSD outperforms existing compared approaches in task scheduling and processor utilization, making it a promising solution for multi-core systems.**

***Keywords*—MultiCore; multiprocessor; DAG scheduling; dynamic priority; task duplication; clustering; MCP**

I. Introduction

With the increasing demand of high-performance and low energy consumption processing, which is a basic requirement in many applications such as image and video processing [1] [2] [3] [4], climate modeling [5] [6] [7], artificial intelligence [8] [9] [10]. Parallel processing is needed to speed up applications performance by splitting its overall job into smaller tasks [11] [12] [13], execute and complete its work across multiple processors. There is an increasing interest in addressing issues related to multi-core chips. The shift to multi-core has emerged because of reaching the physical limits of single core chips, especially clock speed bottleneck [14] [15]. In the last few decades, multi-core processors have evolved from just two cores in a single CPU to multiple cores [11] [12]. It is challenging to find a computer with single-core CPU, as even low-power CPUs are now designed with two or more core per chip [16]. Intel already launches Intel® Xeon® 6 server processor, code-named Sierra Forest with up to 144 cores [17].

In traditional multiple processors, scheduling problem appears to solve the contention between concurrent parts of programs, or arrange programs execution to guarantee enhancement in the overall performance [13]. In multi-core processors the scheduling problem gets worse with the presence of many cores, where a program can be seen as a set of tasks which can run serially or parallelly. The relationship among these tasks may or may not include precedence constrains [11] [12]. Precedence constrains indicates when one task begins or ends in relation to another task [13]. If precedence constrains exist, a Directed Acyclic Graph (DAG) is used to build a task model [12] [13] [18].

The presence and use of multi-core processors is more popular than the traditional multiple processors. In multi-core processors, the scheduling problem is magnified, which is a non-deterministic polynomial (NP-complete) problem [19] [20]. Most of the traditional parallel processing algorithms are designed to handle one-core processors or designed for multi-core processor, but it does not fit multi-core multi-processors [18]. The primary goal of all these algorithms is to try to reduce the program execution time. The DAG based Heuristic algorithms, which can be divided into four categories, which are List based task scheduling algorithms [21] [22] [23] [24] [25] [26], Task Duplication-based scheduling algorithms [27] [28] [29], Cluster based scheduling algorithms [21] [30] [31], and Multi-Core Oriented scheduling algorithms (MCO), which specialize in the types that deal with multi-core multi-processors machines.

Multi-Core Oriented scheduling algorithms are the focus in this paper. Despite the many advantages and the ongoing manufacturing of multi-core multi-processors systems [17]. MCO scheduling algorithms are not independent types, but they apply concepts and methods from previous types targeting multi-core multi-processors systems (MCMP). This paper introduces some attempts.

The utilization of multi-core processor architecture is growing more common in the realm of high-powered computing. This is considered one of the reasons for the emergence of the MCO scheduling algorithms. Some examples of MCO scheduling algorithms are weighted Earliest Finish Time (wEFT) [32], the Priority Queue Task Duplication scheduling algorithm (PQTD) and Genetic-based Scheduling Algorithm on Multicore (GSAM) [33].

The wEFT [32] algorithm is designed for multi-core processor systems, where it assigns the task with minimum earliest completion time to a certain processor core. wEFT performs well when compared with existing task scheduling algorithms, but wEFT is not the best in average waiting time of the tasks. The PQTD is proposed in [18] for multi-core processors. PQTD uses priority queue and task duplication concepts to map the generated task model to processors. As mentioned in [18] The PQTD algorithm has better performance and better processor utilization compared to TDS [34], and CPFD [35].

The GSAM [33] is try to take advantage of multi-core multi-processors and provided solution for the scheduling problem based on genetic approach. The algorithm is repeatedly executed until it reaches a fixed number of iterations. The GSAM algorithm inherits some defects form GA and multi-cores, such as high computation and time consumption. In addition, GSAM suffers from high power need in multi-core architectures. Existing GA based scheduling algorithms have some disadvantages such as high complexity, high power consumption, poor efficiency, poor processors utilization, etc. [36] [37] [15]. The GA and multi core algorithms are intended for some specific applications and are not suitable for other applications.

The proposed scheduling algorithm combines the qualities of multi-core processors algorithms and multi-processors. The proposed algorithm deals with fine grain task graph applications. It overcomes some problems and combines some of the advantages found in others. It increases the processor utilization and tries to cluster the related tasks to run in the same processor cores to improve the performance. The proposed algorithm first uses static priority for level categorization, and then dynamic priority for tasks in each level while assigning tasks. The proposed algorithm Core Scheduler task Duplication, which is CSD.

The remainder of this paper is organized as follows. The proposed scheduling algorithm is presented in Section II. Section III introduces an application example illustrating the proposed scheduling algorithms steps. Section IV provides the simulation results and discussions. Finally, the conclusion is provided in the last section.

II. The Proposed CSD Scheduling Algorithm Design

A. Task Mode

The properties associated with any parallel program such as processing time, data dependencies, communication cost, and synchronization requirements must be known before scheduling. The parallel program is represented by node and edge DAG [13].The DAG task model G = {T, E, C, W} where the set of nodes T is a set of n tasks, where n is the total number of tasks. E is a set of edges in the DAG between two vertices (nodes) Ti, and Tj,where $0 \leq$ i,j $\leq$ n. Cij is as set of communication time between two tasks, where $0 \leq$ i,j $\leq$ n. W is a set of processing time of each task. Fig. 1 is an example of DAG with four tasks numbered. from T1 to T4, T1 has processing time equal to 2 unit of time. T1 and T2 are connected with an edge with communication weight equal to 1 unit of time.

B. Assumptions and Constraints

Some restrictions are necessary to explain the CSD scheduling algorithm, which are as follows:

- CSD targets a machine with a set of n homogenous Processing Elements (PE), where each PE contains a set of m homogenous cores idiomatically called (n) multiprocessor (m) multi-cores system. The core j inside Processing Element i is called (PEi-ci).

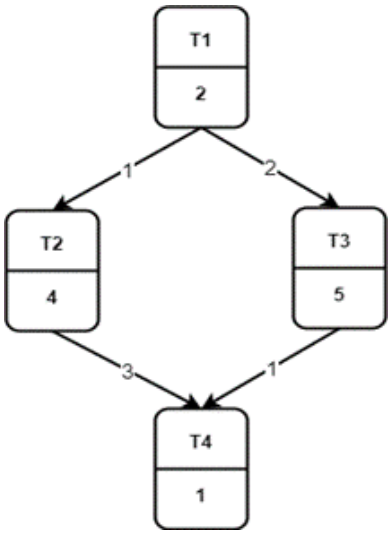


Fig. 1. Example of Directed Acyclic Graph (DAG).

- The processors are fully connected, where link contention and routing strategies used for communication are neglected.
- The communication delay Cij between two cores in the same processor element is negligible (Cij $\approx$ 0), while it takes cij units of time if the two cores belong to different processor elements.
- Any task cannot be started until all its predecessors are completed.
- Assignment procedure will be initiated if either any core is free or the task has been completed, which leads to a free core.
- Transition from any level to another is not allowed until the assignment of all tasks at this level are completed.

C. CSD Scheduling Algorithm Procedure

CSD scheduling algorithm is formalized in this section.

Input:

- A DAG describes tasks workloads, their weights, communication cost between any tasks, and precedence relation.
- The number of PE and how many cores each one contains.

Output: Gantt chart illustrates the list of all tasks allocated to each core ordered by their starting, finishing, and execution time. Gantt chart also indicated Idle periods for every core.

Steps: Fig. 2 shows the steps of CSD algorithm, while the part concerned with selecting task from unassigned level list is illustrated in Fig. 3. These steps are listed as follows:

1) Arrange the given M cores of each processing elements PE lexicographically in a queue.

2) Build the static task queue, where this queue is divided into static priority levels starting from Level1 (L1) to Levelm (Lm). Tasks belonging to any level Li only depend on the tasks belonging to the previous level Li-1. Tasks arranged at this point have no order. Only the tasks belonging to the same level are stacked together with no priority. Each level is rearranged during the assignment procedure step and accordingly the priorities of the tasks are changed during the successive stages of CSD algorithm within the same level. Because of this change, the task static priorities turn to dynamic priorities.
3) Collect and compute all the following characteristics for each task before the assignment procedure:
   - weight of the task i, $0 \leq i \leq n$ (Wi),
   - list of the predecessors of the task, i.e. parents (predi),
   - successors of the task, i.e. children (succi),
   - the task should be duplicated or not and how many times it should be duplicated (D) according to Eq. (1).

$$\min(\#ofPEs, ceil(\frac{\#ofchildren}{\#ofcores})) \quad (1)$$

   As the number of tasks in the same level increases, the congestion problem emerges and worsens with the duplication according to the (1). This problem was resolved using Eq. (2).

$$\min(\#ofPEs, floor(\frac{\#ofchildren}{\#ofcores})/\#oftasks) \quad (2)$$

   - The task found in critical path (CP) in the DAG, which is the longest path in the DAG or not,
   - degree of the CP, which is number of tasks in the CP, the more tasks the higher the degree (DCP), and located on more than one CP.
   - Select the tasks that can be duplicated if possible and mark it as duplicated task. The selection process satisfies the inequality: # of children $>$ # of cores.
   - Any task chosen for duplication is indeed duplicated in the task queue many times as Eq. (1) or Eq. (2).
4) Assign previously arranged tasks to cores: this step is called assignment procedure, which assigns the tasks to cores and is carried out periodically in every time slot. The procedure is divided into two phases. The first phase is called **core phase** and deals with the cores. Meanwhile the second phase is called **task phase** and deals with the selection of the task to be performed. Each phase is explained as follows.
   Core phase: Roll over all free cores to see if they can start any selected task from the next phase. It should be noted that the list of free cores changes every time slot, where the algorithm picks all free cores. The CSD algorithm determines, which core suitable for the priority based ordered tasks.
   Task phase: Priorities change at this phase based on a number of criteria to choose the task, which has the highest priority. These priorities change periodically with each start of task selection. CSD considers all the tasks in the same level and rearrange them again, then choose the task that will be executed if:
   • One of its predecessors is completed, and none of its duplicated is running in the same processing element PE. If a tie occurs, then go to the next step.
   • When CSD algorithm reaches this step, there is more than one choice, which are:
   First: There is only one task in CP,
   Second: There are two or more tasks in CP.
   Third: no task in CP.
   For the first case the algorithm chooses this task to be scheduled. For the second and third track, the algorithm calculates the longest path of Ti where the task is on (LPi). The LPi can be calculated using Eq. (3):

$$LPi = longestpathbefore + longestpathafter - Wi \quad (3)$$

   for each task, if two or more tasks have the same LPi length, then the algorithm calculates Forest Cost of task i (FCi). FCi is the number of edges in the following subgraph for this task. If tie occurs again then the algorithm moves to the next action.
   - Choose the task with greater number of children if tie occurs then.
   - Choose the one with lower weight, if tie occurs then.
   - Choose task with small index.

   After selecting the task, the CSD determines if the task is duplicated and how many times. CSD has to decide whether to start all the duplicated tasks all at the same time at available free cores and remove the exceeded copies from task queue. The other choice for CSD is to start the duplicated task any time as there are free cores.
   Repeat step 4 till each free core from core phase gains a task. Consequently, the same steps are repeated and moved from one level to another until the assignment procedure is completed for all tasks.

III. An Application Example

The following example illustrates how CSD algorithm works.

Assume that the DAG shown in Fig. 4 is given. The objective is to schedule this DAG to 2 PE with 2 cores each. The DAG has 10 nodes, 14 edges, and with CP equal to 13. In the Fig. 4, CP is shown with thick arrows.

1) Arrange the PE, with its cores in processor queue as shown in Fig. 5.
2) Arrange tasks in static levels as shown in Fig. 5.
3) collect all the information about each task in advance as Shown in Table I.
4) Assignment procedure:
   - L1 has only one node (entry node) which is marked as duplicated (2-times) on different cores (PE0-c0), and (PE1-c0).

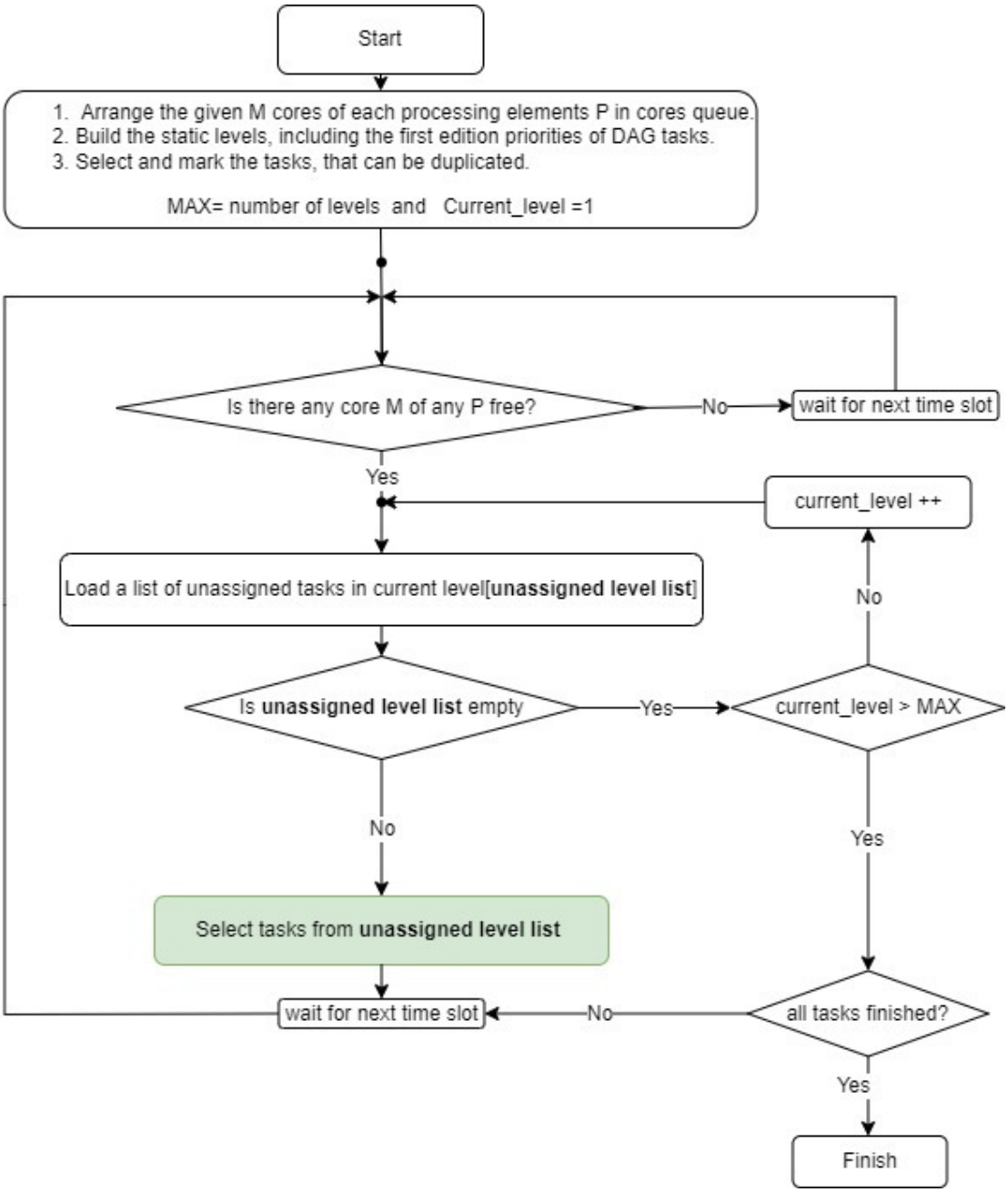


Fig. 2. CSD Scheduling algorithm flowchart.

- The assignment procedure suspended till T1 finished.
- CSD algorithm dynamically rearranges tasks in L2 to get ready_list1, followed by ready_list2: Ready_list1: T2, T3, T4, T5
  Ready_list2: T2, T3, T4, T5
  Final arrangement in this step: T3, T2, T4, T5
- CSD algorithm assigns T3 to (PE0-c0). At the same time slot, it assigns T2 to (PE0-c1), T4 to (PE1-c0), and T5 to (PE1-c1). T4 is marked as duplicated, but there are no free cores available for duplication. At this moment all cores are busy, so CSD waits until one becomes free. The beginning of time slot 4 (PE0-c0), and (PE1-c0) will be free.
- At this moment when L2 is finished CSD algorithm goes to L3, but there are no tasks ready. Therefore, the algorithm waits until the next time slot.
- The algorithm rearranges tasks in L3 until it reaches the final arrangement according to the illustrated criteria in the algorithm procedure. The Final arrangement became T6, T7, T8.
- The algorithm assigns T6 to (PE0-c0).
- T7 is assigned to any of the following (PE0-c1) or (PE1-c0) or (PE1-c1), where (PE0-c1)

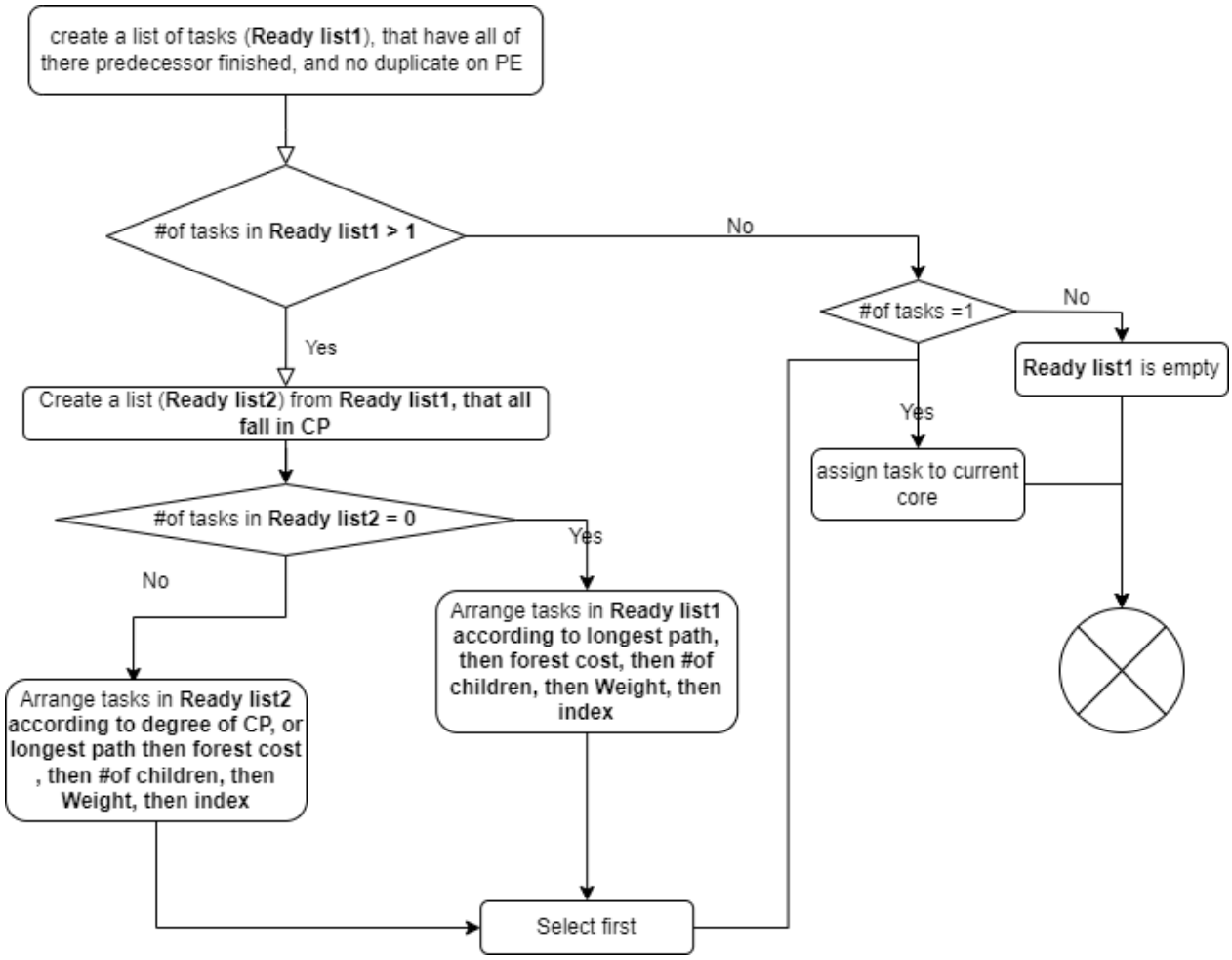


Fig. 3. Task selection flowchart.

TABLE I. TASK CHARACTERISTICS

	T1	T6
ID	1	6
Wi	1	2
Duplicated	yes	no
D	2	2
predi	2	2
succi	–	2
found in CP or not	yes	yes
DCP	5	5

starts after 2 time slots from time slot 4, and (PE1-c0) or (PE1-c1) start immediately, so assign T7 to (PE1-c0).

- T8 has similar situations as T7, so the algorithm assigns T8 to (PE1-c1).
- (PE0-c1) is free, and L3 is finished, but T9 cannot start until T6 is finished, so the algorithm waits.
- When T6, T7, and T8 are also finished, then all cores are free. However, the most suitable core is (PE0-c0), so the algorithm assigns T9 to (PE0-c0).
- T10 cannot start until all predecessors are finished (T9, T7, and T8). Once finished, all cores become free. Either PE0 or PE1 starts the execution of T10 at the beginning of the 8th time slot, so all cores have the same chance to execute it, so the algorithm assigns T10 to (PE0-c0).
- The output of the previous steps represented in Gantt chart shown in Fig. 6. The Gantt chart represents the Schedule Length (SL). SL is the total duration or makespan of the schedule, which emphasizes all the tasks is completed.

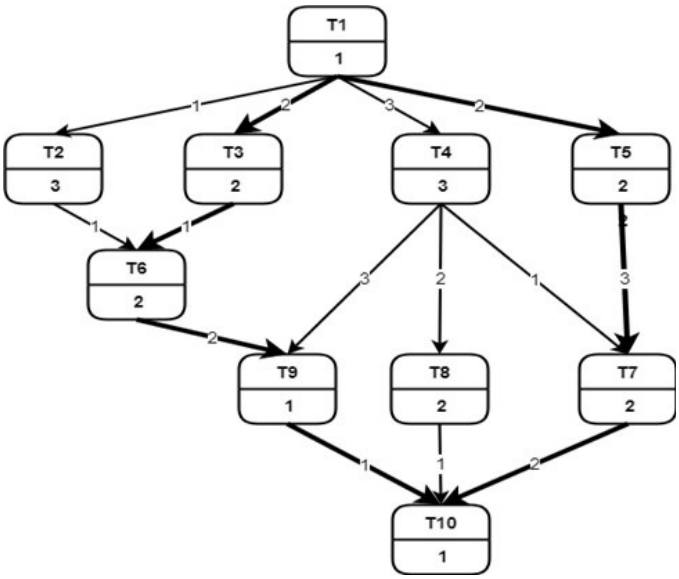


Fig. 4. A sample DAG from [18].

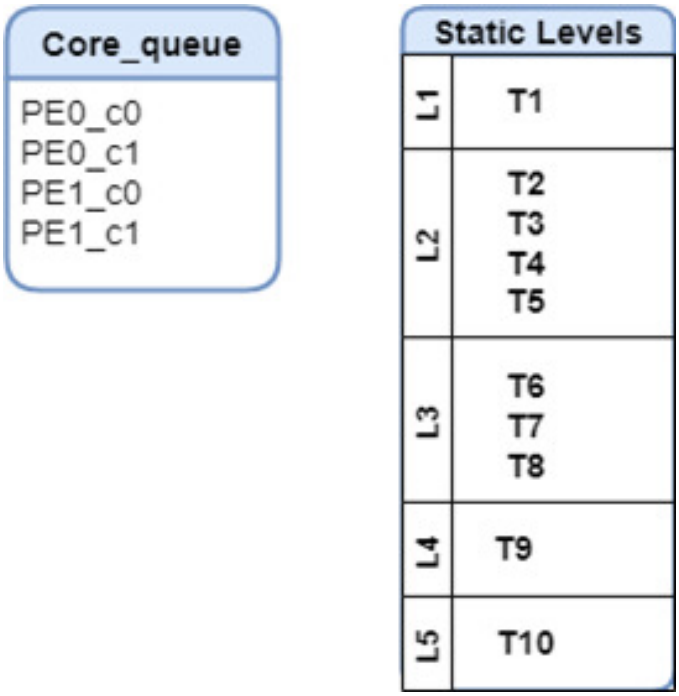


Fig. 5. Arrangement of core queue, and static task levels.

IV. Simulation and Results

The performance of the proposed scheduling algorithm is evaluated using randomly generated task graphs or task graphs modeled from actual application programs from Standard Task Graph set, which is presented in [38].

1) The first task graph t50_rand1 is composed of 50 tasks and 985 edges with computation to communication ratio (CCR) equal 0.1.
2) t100_rand2 composed of 100 tasks and 3943 edges with CCR equal 0.02.

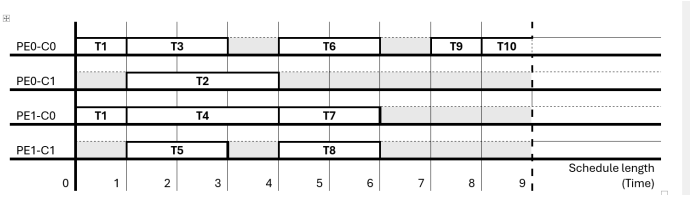


Fig. 6. The schedule generated by CSD algorithm.

3) and robot control application with 88 tasks and 131 edges with CCR equal 0.85 and 5.

The CSD scheduling algorithm simulation is conducted using Python running several times with various processing elements / cores settings. Fig. 7 presents the simulation results of the CSD scheduling algorithm with varying numbers of Processing Elements (PEs), each containing only one core. The x-axis represents the number of PEs, while the y-axis indicates the Schedule Length (SL). The figure also compares the performance of different versions of the CSD scheduling algorithm with the Modified Critical Path (MCP) scheduling algorithm. The difference between the multiple versions of CSD is when to duplicate the chosen task to be duplicated and how many times:

- If duplication occurs during the first copy assignment process at the same level according to (1) Version CSD (once_nolvl) is generated.
- If duplication occurs any time according to (1) Version CSD (any_nolvl) is generated.
- If duplication occurs during the first copy assignment process at the same level according to (2) Version CSD (once_lvl) is generated.
- If duplication occurs any time according to (2) Version CSD (any_lvl) is generated.
- CSD (none) is generated when no duplication occurs.

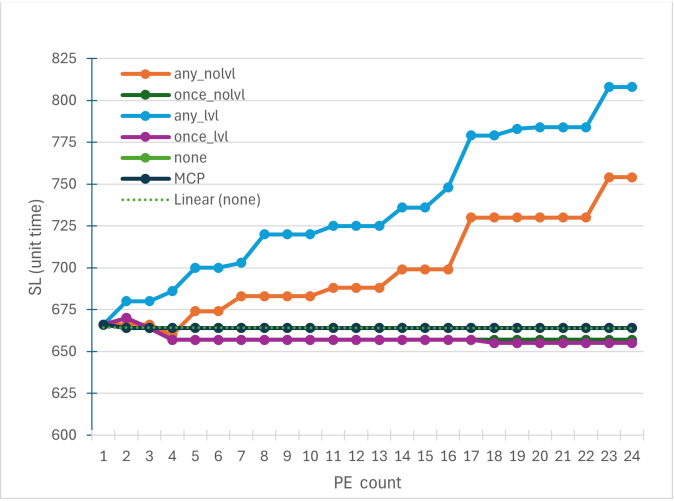


Fig. 7. Scheduling for t50_rand1 DAG.

As seen in Fig. 7 the SL is improved with the CSD (once_lvl and once_nolvl) versions compared to other algo-

rithms. CSD (once_lvl) outperforms CSD (once_nolvl) when the number of tasks in the same level increases. The performance of CSD (none) algorithm coincides with MCP algorithm. Both CSD (none) and MCP algorithms outperform all the reset versions of CSD. While the CSD (any_nolvl) and (any_lvl) deviated too far as it exhausted the free cores in duplication while there are already tasks to begin execution. This scenario is reflected in a significant increase in SL. CSD(Any_nolvl) and CSD(any_lvl) and the results are ignored. Fig. 8 illustrates SL for t100_rand2, which is like t50_rand1. It is noticeable that there is an improvement of CSD (once_lvl) over MCP, as the number of parallel paths increase. Fig. 9 illustrates the average SL output behavior with robot control application generated with CCR equal 0.85 and 5. Robot control DAG in contrast to other DAGs shows a significant improvement performance of MCP over all version of the proposed CSD. This is due to the decrease in parallel paths, which is the primary motivation and cause for duplication.

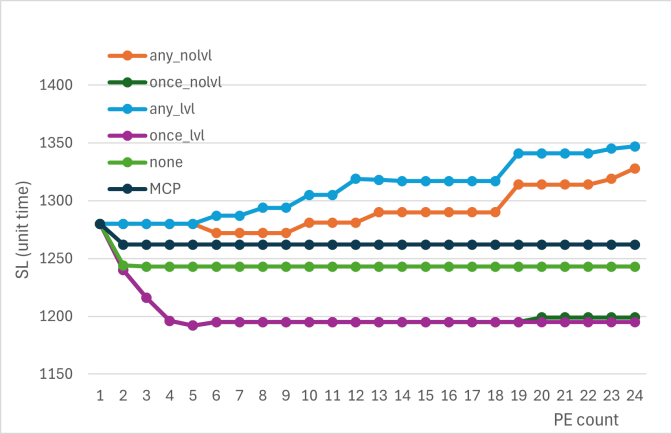


Fig. 8. Scheduling for t100_rand2 DAG.

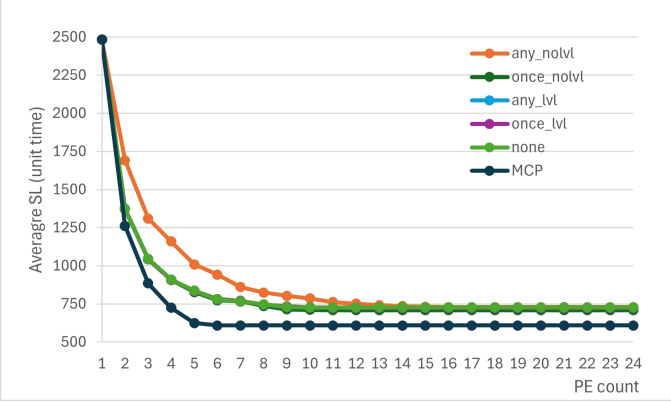


Fig. 9. Average SL for robot control with (0.85,5) CCR.

The CSD scheduling algorithm is simulated again with one PE only that has varying number of cores. Fig. 10 shows the SL with robot control DAG described before. The chosen DAG indicates almost the same response as robot task graph when applying this PE/Cores configuration. The comparisons here are between the versions of CSD and the optimal SL. Optimal SL is the length of CP without communication. The primary goal of the CSD algorithm is to reduce SL by clustering the tasks related to others and give higher priority to CP tasks. CSD algorithm closely approach the optimal demonstrating its effectiveness in finding near-optimal SL.

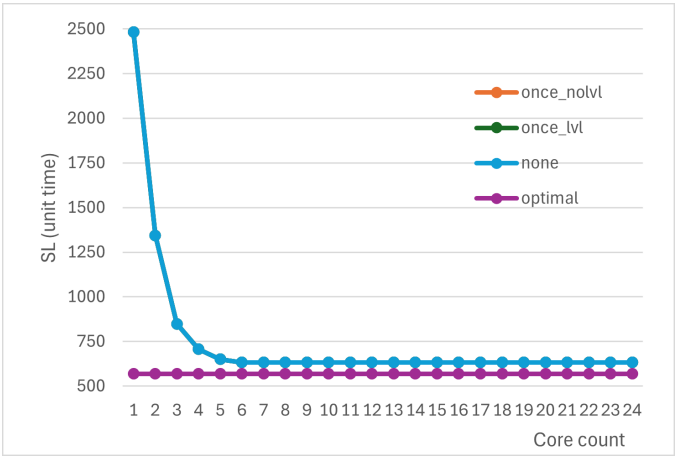


Fig. 10. Scheduling for robot control 88 tasks DAG.

Fig. 11 shows the result when the CSD algorithm is simulated with different number of PEs that have varying number of cores. Given that the total number of processing units (i.e. PE or core) is fixed and equal to 24. xxx shows the average SL for CSD versions on all the task graphs with different CCR. CSD algorithm produces schedule close to optimal as the number of cores increase.

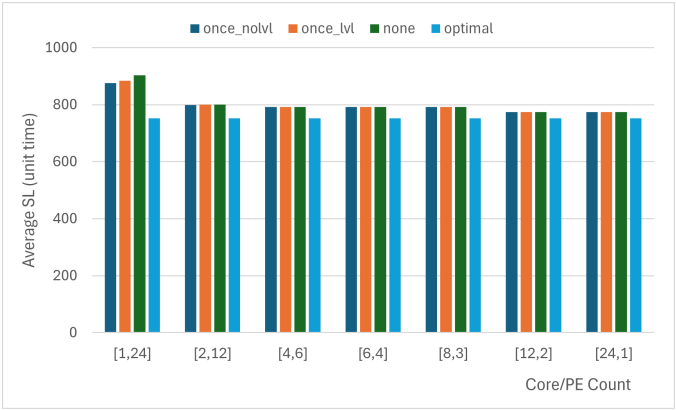


Fig. 11. Average SL for all task graphs.

V. Conclusion

This paper proposed multiple versions of a scheduling algorithm called CSD. Each version introduces enhancements and modifications that change with the workload behavior. CSD in its primitive version with no duplication matches MCP in the case the used architecture has only one core with different PEs. When comparing the (once_nolvl) version of CSD with MCP, the proposed algorithm outperforms MCP in SL, throughput and processors utilization. The results indicate that no single algorithm outperforms others in all scenarios. Instead, the effectiveness of a scheduling algorithm is dependent on factors

such as the nature of workload and architecture of multi-processor/multi-core system.

REFERENCES

[1] A. Kika and S. Greca, "Multithreading image processing in single-core and multi-core cpu using java," *International Journal of advanced computer science and applications*, vol. 4, no. 9, 2013.

[2] K. M. Hosny, A. Salah, and A. Magdi, "Parallel image processing applications using raspberry pi," in *Recent Advances in Computer Vision Applications Using Parallel Processing*. Springer, 2023, pp. 107–119.

[3] A. Kamalakannan and G. Rajamanickam, "High performance color image processing in multicore cpu using mfc multithreading," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 12, pp. 42–47, 2013.

[4] K. Mia, T. Islam, M. Assaduzzaman, T. M. N. U. Akhund, A. Saha, S. P. Shaha, M. A. Razzak, and A. Dhar, "Parallelizing image processing algorithms for face recognition on multicore platforms," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 11, 2022.

[5] T. Radhika, K. Gouda, and S. S. Kumar, "Novel approach for spatiotemporal weather data analysis," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 7, 2022.

[6] L. Su and S. Naffziger, "1.1 innovation for the next decade of compute efficiency," in *2023 IEEE International Solid-State Circuits Conference (ISSCC)*. IEEE, 2023, pp. 8–12.

[7] J. Subha and S. Saudia, "Integrating regression models and climatological data for improved precipitation forecast in southern india," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 5, 2023.

[8] M. N. Al-Andoli, K. S. Sim, S. C. Tan, P. Y. Goh, and C. P. Lim, "An ensemble-based parallel deep learning classifier with pso-bp optimization for malware detection," *IEEE Access*, 2023.

[9] R. Pirayeshshirazinezhad, S. G. Biedroń, J. A. D. Cruz, S. S. Güitrón, and M. Martínez-Ramón, "Designing monte carlo simulation and an optimal machine learning to optimize and model space missions," *IEEE Access*, vol. 10, 2022.

[10] R. A. Jain and D. V. Padole, "Scalable and flexible heterogeneous multi-core system," *International Journal of Advanced Computer Science and Applications*, vol. 3, no. 12, 2012.

[11] G. S. Almasi and A. Gottlieb, *Highly parallel computing*. Benjamin-Cummings Publishing Co., Inc., 1994.

[12] T. Rauber and G. Rünger, "Parallel programming: For multicore and cluster systems," *Citado na*, p. 30, 2013.

[13] H. El-Rewini and M. Abd-El-Barr, *Advanced computer architecture and parallel processing*. John Wiley & Sons, 2005.

[14] H. Esmaeilzadeh, E. Blem, R. St. Amant, K. Sankaralingam, and D. Burger, "Dark silicon and the end of multicore scaling," in *Proceedings of the 38th annual international symposium on Computer architecture*, 2011, pp. 365–376.

[15] M. D. Hill and M. R. Marty, "Amdahl's law in the multicore era," *Computer*, vol. 41, no. 7, pp. 33–38, 2008.

[16] M. Gupta, L. Bhargava, and S. Indu, "Mapping techniques in multicore processors: current and future trends," *The Journal of Supercomputing*, vol. 77, no. 8, pp. 9308–9363, 2021.

[17] I. Website, "Intel." [Online]. Available: https://www.intel.com/content/www/us/en/products/details/processors.html

[18] X. Yao, P. Geng, and X. Du, "A task scheduling algorithm for multi-core processors," in *2013 International Conference on Parallel and Distributed Computing, Applications and Technologies*. IEEE, 2013, pp. 259–264.

[19] X. Xiao and Z. Li, "Chemical reaction multi-objective optimization for cloud task dag scheduling," *IEEE Access*, vol. 7, pp. 102 598–102 605, 2019.

[20] T. Lively, W. Long, and A. Pagnoni, "Analyzing branch-and-bound algorithms for the multiprocessor scheduling problem," *arXiv preprint arXiv:1901.07070*, 2019.

[21] M.-Y. Wu and D. D. Gajski, "Hypertool: A programming aid for message-passing systems," *IEEE transactions on parallel and distributed systems*, vol. 1, no. 3, pp. 330–343, 1990.

[22] H. Topcuoglu, S. Hariri, and M.-Y. Wu, "Task scheduling algorithms for heterogeneous processors," in *Proceedings. Eighth Heterogeneous Computing Workshop (HCW'99)*. IEEE, 1999, pp. 3–14.

[23] S. Branch and I. Shoushtar, "List-scheduling techniques in homogeneous multiprocessor environments: a survey," *International Journal of Software Engineering and Its Applications*, vol. 9, no. 4, pp. 123–132, 2015.

[24] G. C. Sih and E. A. Lee, "A compile-time scheduling heuristic for interconnection-constrained heterogeneous processor architectures," *IEEE transactions on Parallel and Distributed systems*, vol. 4, no. 2, pp. 175–187, 1993.

[25] H. El-Rewini and T. G. Lewis, "Scheduling parallel program tasks onto arbitrary target machines," *Journal of parallel and Distributed Computing*, vol. 9, no. 2, pp. 138–153, 1990.

[26] Y.-K. Kwok and I. Ahmad, "Dynamic critical-path scheduling: An effective technique for allocating task graphs to multiprocessors," *IEEE transactions on parallel and distributed systems*, vol. 7, no. 5, pp. 506–521, 1996.

[27] I. Ahmad and Y.-K. K. Y.-K. Kwok, "A new approach to scheduling parallel programs using task duplication," in *1994 Internatonal Conference on Parallel Processing Vol. 2*, vol. 2. IEEE, 1994, pp. 47–51.

[28] Y.-C. Chung *et al.*, "Applications and performance analysis of a compile-time optimization approach for list scheduling algorithms on distributed memory multiprocessors," in *SC Conference*. IEEE Computer Society, 1992, pp. 512–521.

[29] G.-L. Park, B. Shirazi, and J. Marquis, "Dfrn: A new approach for duplication based scheduling for distributed memory multiprocessor systems," in *Proceedings 11th international parallel processing symposium*. IEEE, 1997, pp. 157–166.

[30] J.-J. Hwang, Y.-C. Chow, F. D. Anger, and C.-Y. Lee, "Scheduling precedence graphs in systems with interprocessor communication times," *siam journal on computing*, vol. 18, no. 2, pp. 244–257, 1989.

[31] S. Cao and J. Bian, "Improved dag tasks stretching algorithm based on multi-core processors," in *2020 IEEE 11th International Conference on Software Engineering and Service Science (ICSESS)*. IEEE, 2020, pp. 18–21.

[32] L. Liu and D. Qi, "An independent task scheduling algorithm in heterogeneous multi-core processor environment," in *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*. IEEE, 2018, pp. 142–146.

[33] J. Pecero, S. Varrette, and P. Bouvry, "Scheduling dag applications on multi-core processor packages architectures," 2010.

[34] S. Darbha and D. P. Agrawal, "Optimal scheduling algorithm for distributed-memory machines," *IEEE transactions on parallel and distributed systems*, vol. 9, no. 1, pp. 87–95, 1998.

[35] I. Ahmad and Y.-K. Kwok, "On exploiting task duplication in parallel program scheduling," *IEEE Transactions on parallel and distributed systems*, vol. 9, no. 9, pp. 872–892, 1998.

[36] R. Medina, E. Borde, and L. Pautet, "Scheduling multi-periodic mixed-criticality dags on multi-core architectures," in *2018 IEEE Real-Time Systems Symposium (RTSS)*. IEEE, 2018, pp. 254–264.

[37] S. Shah, A. Qahir, M. Safeer, S. Mazahir, and O. Hasan, "Comfast: A comparative framework for analysis of scheduling techniques in multi-core systems," in *2018 Annual IEEE International Systems Conference (SysCon)*. IEEE, 2018, pp. 1–7.

[38] Kasahara; H.; Tobita; T.; Matsuzawa; Sakaida, "S.: Standard task graph set." [Online]. Available: https://www.kasahara.cs.waseda.ac.jp/schedule/

Enhancing Skin Cancer Detection with Transfer Learning and Vision Transformers

Istiak Ahmad, Bassma Saleh Alsulami, Fahad Alqurashi
Department of Computer Science-Faculty of Computing and Information Technology,
King Abdulaziz University, Jeddah 21589, Saudi Arabia

***Abstract*—Early and accurate detection of skin cancer is critical for effective treatment. This research aims to enhance skin cancer multi-class classification using transfer learning and Vision Transformers (ViTs), addressing the challenges of imbalanced medical imaging data. We introduced data augmentation techniques to the HAM10000 dataset to enhance the diversity of the training and implemented 13 pre-trained transfer learning models. These included DenseNet (121, 169, and 201), ResNet (50V2, 101V2, and 152V2), VGG (16 and 19), NasNet (mobile and large), InceptionV3, MobileNetV2, and InceptionResNetV2, as well as two Vision Transformer architectures (ViT and deepViT). After fine-tuning these models, DenseNet121 achieved the highest accuracy of 94%, while deepViT reached 92%, highlighting the effectiveness of these approaches in skin cancer detection. Future work will focus on refining these models, exploring hybrid approaches that combine convolutional neural networks and transformers, and expanding the framework to other cancer types to advance automated diagnostic tools in dermatology.**

***Keywords*—Medical imaging; skin cancer; multi-class classification; detection; deep learning; transfer learning; vision transformer**

I. Introduction

Skin cancer [1] typically starts in the skin cells when skin cells are damaged, usually from too much exposure to the sun's ultraviolet (UV) rays. The three main types of skin cancer are basal cell carcinoma, squamous cell carcinoma, and melanoma, where melanoma is the most serious. People with fair skin, a history of sunburns, prolonged sun exposure, a family history of skin cancer, or a weakened immune system are at the highest risk. Symptoms of skin cancer include the appearance of new or unusual skin growths, changes in the size, shape, or colour of moles, and persistent non-healing sores. According to the American Cancer Society [2], in 2024, there are estimated to be 100,640 new cases of melanoma skin cancer (59,170 male and 41,740 female) and an estimated 8,290 deaths (5,430 male and 2,860 female). The National Cancer Institute [3] reported that the relative survival rate for skin cancer over the five years from 2014 to 2020 is 94.1%. Early skin cancer detection is crucial, as it significantly enhances the likelihood of effective therapy and survival.

Artificial intelligence plays a pivotal role in enhancing sustainability across various sectors, including education [4], transportation [5], cybersecurity [6], [7], social media [8], [9] and healthcare. Notably, the field of healthcare has experienced remarkable advancements, particularly in the early detection of cancer, with AI's contributions proving to be increasingly impactful. The prompt identification of skin cancer enables using less invasive treatment approaches, reducing the risk of severe consequences and improving the patient's prognosis. The process of developing an early skin cancer detection system involves integrating cutting-edge imaging technology with advanced deep-learning algorithms. These appliances have the capability to accurately analyse skin lesions, detecting distinguishing patterns that may signify the presence of cancer. Computer-aided diagnosis (CAD) is a vital component of these systems, providing dermatologists with a valuable tool. CAD systems enhance diagnostic accuracy by analyzing visual data from skin images and distinguishing between benign and malignant tumours. CAD systems are essential for improving skin cancer's early and accurate diagnosis [10], improving patient outcomes by providing reliable views, and eliminating diagnostic errors. Skin cancer detection encounters challenges, including limited, diverse datasets, which hinder model generalization and class imbalance, leading to biased predictions. The visual similarity between benign and malignant lesions complicates detection, while the "black box" nature of AI models raises interpretability issues. Overcoming these hurdles is essential for enhancing skin cancer detection technologies. Transfer learning and Vision Transformers (ViTs) enhance model performance and generalization to overcome challenges with skin cancer detection. Transfer learning utilizes pre-trained models on extensive datasets to improve accuracy, even when skin cancer data is limited, and to address the issue of imbalanced classes. Vision Transformers (ViTs) improve the distinction between visually similar benign and malignant tumours by analyzing complex patterns and establishing connections across long distances in images.

A. The Aim and Objectives

The aim of this research is to develop and evaluate a robust multiclass skin cancer detection system by leveraging transfer learning, Vision Transformers, and advanced data augmentation techniques to achieve high accuracy. The research contributions are as follows:

- Applied refined data augmentation techniques to address challenges of limited and imbalanced datasets, significantly improving model robustness and generalization.
- Conducted a comprehensive analysis of 13 transfer learning models, underlining their comparative performance and demonstrating significant improvements in skin cancer detection accuracy.
- Implemented two Vision Transformer models, effectively leveraging their ability to capture intricate patterns in skin lesion images, enhancing detection accuracy.

The remainder of the paper is organised as follows: Section II discusses the literature review and research gaps. Section III discusses the proposed research methodology. Section IV discusses the research outcomes for transfer learning models and vision transformers models. Section V compares the overall research outputs with existing research and discusses research challenges and limitations. Finally, Section VI concludes by stating the future direction.

II. Literature Review

Kondaveeti et al. [11] proposed pre-trained models, such as ResNet50, MobileNet, Xception, and InceptionV3 for 7 types of skin cancer detection. They achieved the highest 90% accuracy, 89% weighted average precision, and 90% recall for ResNet50 on the HAM10000 dataset. Naik et al. [12] proposed MobileNetV2 for skin cancer detection and achieved 93.11% accuracy. Fraiwan and Faouri [13] used 13 transfer learning models to detect 7 types of skin cancer. They achieved the highest 82.9% accuracy for the DenseNet201 model. Swetha et al. [14] compared several transfer learning models, such as ResNet50, ResNet152, ResNet101, VGG16, VGG19, MobileNet, and Xception for multiclass skin cancer detection and obtained 83.69% categorical accuracy. Vishnu et al. [15] presented augmentation techniques and ensemble models by integrating InceptionV3 and DenseNet201 weights and outputs for six types of skin cancer detection. They achieved 89% accuracy with a validation loss of 0.44. Kaveti et al. [16] proposed a ResNet101 model for multiclass skin cancer detection using the HAM1000 dataset and got an accuracy of 92% for seven skin cancer categories. Arshed et al. [17] employed a pre-trained vision transformer model for seven types of skin cancer detection and achieved 92.14% accuracy. Tuncer et al. [18] presented lightweight CNN-based techniques for detecting benign and malignant types of skin cancer. They achieved 92.12% accuracy using the HAM10000 dataset. Yang et al. [19] proposed attention-weighted transformers for skin cancer detection using the HAM10000 dataset and achieved 93.75% accuracy. Ashfaq and Ahmad [20] implemented InceptionResNetV2, VGG16, ResNet50, EfficientNetB3, and vision transformer B32 for multiclass skin cancer classification. Among these models, EfficientNetB3 obtained the highest accuracy of 86% with precision 85%, recall 79%, and F1-score 81%. Sanchez et al. [21] proposed EfficientNet (B0, B1, B2, and B3) and ViT (base 16, base 32, large 16 and large 32) models and obtained the highest accuracy of 85% by fine-tuning the number of epochs and learning rate.

Despite considerable progress in skin cancer detection, key gaps remain that align with this study's objectives. Existing research has explored data augmentation to address imbalanced datasets; however, more refined techniques are needed to further enhance model robustness. This study addresses the scarcity of comprehensive comparisons of multiple transfer learning models' performances by evaluating 13 models, thereby providing a clearer insight into their effectiveness. Moreover, Vision Transformers (ViTs) have shown promise, but their potential to capture detailed patterns in skin lesion images remains underexplored. This research advances the field by implementing and testing two ViT models, demonstrating their capability to enhance detection accuracy.

III. Methodology and Design

Fig. 1 presents the proposed research methodology for developing and evaluating models for cancer detection using transfer learning and vision transformer techniques. The methodology is divided into several key steps. The first step includes dataset collection from the HAM10000 dataset. We have conducted several exploratory data analyses, including analysing the distribution of classes and disease distribution across genders, to acquire an understanding of the dataset. We have discovered that the dataset is imbalanced, which can lead to bias and overfitting issues. Data augmentation techniques are implemented to resolve these issues and improve model generalization. The more details of the dataset distribution are discussed in Section III-A. This study implemented several transfer learning models (see Section III-B) and vision transformer models (see Section III-C) for cancer detection. Finally, the models are evaluated using evaluation metrics (see Section III-D).

Algorithm 1 shows the process of our proposed methodology. Initially, we read all the images from the directory and saved them with the corresponding labels (line numbers 23 and 24). After that, image preprocessing is employed to remove noise from the image, and image augmentation techniques are applied to balance the number of images for each class (lines 25 and 26). The function AUGMENTEDPIPELINE discusses the applied augmentation techniques in this study (line numbers 1 to 11). Following this, the dataset is split into training, validation, and testing sets equally (line number 27). Then, the Keras transfer learning pre-trained model is loaded as a base model and added to a fully connected layer (FCmodel) (line numbers 29 and 30). The detailed architecture of a fully connected layer is given from lines numbers 12 to 22. Furthermore, the FCmodel, Adam optimizer, and entropy loss function are passed as parameters to the compiler (line number 31). Similarly, PyTorch vision transformer modes (DeepViT and ViT) are loaded as transformer models (line number 33). To calculate the training time of the proposed method, a timer is started in line number 34. Sequentially, the compiled model, validation, and training set are passed to train the model (line number 35). After training the model, the model and training history are saved in h5 format (line number 36). Line number 37 provides the model execution time. Finally, we evaluate our proposed method by passing the testing set, saved model, and training history through several evaluation metrics.

A. Dataset

This study used the HAM10000 (Human Against Machine with 10000 training images) [22] dataset, which was presented by the International Skin Imaging Collaboration (ISIC) for skin cancer detection in 2018 [23]. This dataset contains 10,015 dermatoscopic images for seven types of cancer, including Actinic keratoses and intraepithelial carcinoma (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv) and vascular lesions (vasc). Fig. 2 presents the distribution of seven types of skin cancer derived from the HAM10000 dataset. The image IDs, along with the corresponding patient details, such as age, sex, and lesion localization, are as follows: akiec (ISIC_0027930, 60, male, and scalp), bcc (ISIC_0026343, 70, male, and face), bkl (ISIC_0028233, 55, male, and

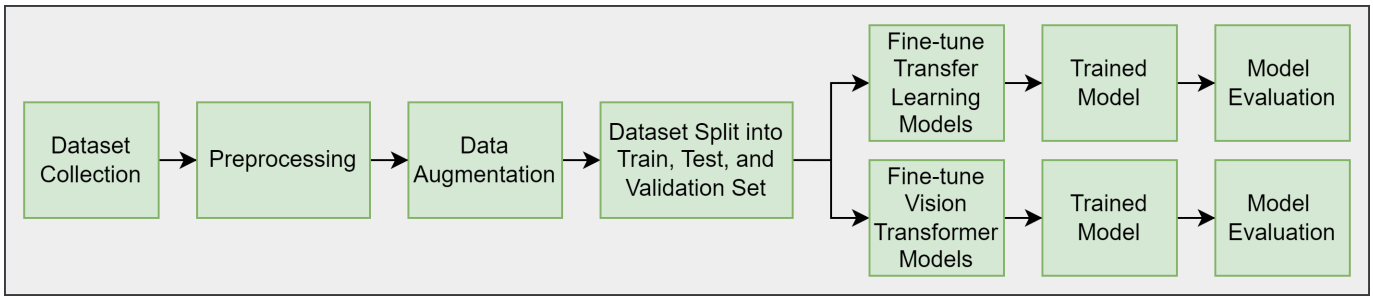


Fig. 1. Research methodology and design.

Algorithm 1 Master Algorithm

Input: *directory path of mages*
Output: *skin cancer classification*

function AUGMENTEDPIPELINE($imgProcess$)
  $imgProcess.Sequential([$
  $imgProcess.Fliplr(0.5),\ \#\ horizontal\ flip$
  $imgProcess.Flipud(0.5),\ \#\ vertical\ flip$
  $imgProcess.Crop(percent = (0,\ 0.1),$
  $imgProcess.MultiplyBrightness(0.8,\ 1.2),$
  $imgProcess.GaussianBlur(sigma = (0, 1.0))),$
  $imgProcess.Grayscale(alpha = (0.0, 1.0)),$
  $imgProcess.CoarseDropout(0.02),$
  $imgProcess.CLAHE(clip\_limit = (1, 4))])$
end function
function FCLAYER($baseModel$)
  $A \leftarrow baseModel.output$
  $A \leftarrow GlobalAveragePooling2D()(A)$
  $A \leftarrow Flatten()(A)$
  $A \leftarrow Dense(128, activation = "relu")(A)$
  $A \leftarrow Dropout(0.3)(A)$
  $A \leftarrow Dense(512, activation = "relu")(A)$
  $A \leftarrow Dropout(0.3)(A)$
  $A \leftarrow Dense(7, activation = "softmax")(A)$
  $FCmodel \leftarrow Model(baseModel.input, A)$
end function
$imageDir \leftarrow directory\ path\ of\ mages$
$imgPathLabel \leftarrow readImages\ (imageDir)$
$imgProcess \leftarrow imagePreprocessing\ (imgPathLabel)$
$imgAugment \leftarrow augmentedPipeline\ (imgProcess)$
$train, test, valid \leftarrow dataset\ split\ (imgAugment)$
$\#\ Transfer\ Learning\ Models:\ using\ Tensorflow, Keras$
$baseModel \leftarrow pretrainedModel(model)$
$FCmodel \leftarrow FClayer(baseModel)$
$Model \leftarrow compileModel(FCmodel, optimizer = adam, loss = entropy)$
$\#\ Transformer\ Models:\ using\ PyTorch$
$Model \leftarrow transformer(baseModel, parameters, optimizer = adam, loss = entropy)$
$start \leftarrow time.time()$
$Fmodel \leftarrow trainModel(Model, train, validation)$
$savedModel, history \leftarrow save(Fmodel)$
$executeTime \leftarrow time.time() - start$
$evalMetrics \leftarrow evaluation(test, savedModel, history)$

face), df (ISIC_0028880, 55, male, and lower extremity), nv (ISIC_0028888, 50, male, and trunk), mel (ISIC_0029241, 70, male, and face), and vasc (ISIC_0027790, 50, female, and face).

Table I provides a comprehensive overview of the HAM10000 dataset, demonstrating the distribution of images across different classes and dataset splits. Each class includes a specific number of original and augmented images, as well as their distribution across training, validation, and test sets. As the original dataset is imbalanced, we implemented data augmentation techniques such as horizontal and vertical flipping, cropping, rotation, adjusting brightness and contrast, applying Gaussian blur, and adding Gaussian noise to address the class imbalance and improve the model's generalization capability.

TABLE I. DISTRIBUTION OF ORIGINAL, AUGMENTED, AND SPLIT DATASET IMAGES ACROSS SEVEN DIAGNOSTIC CLASSES

Classes	Original Images	Augmented Images	Train	Valid	Test
akiec	327	6705	3747	961	1997
bcc	514	6705	3710	958	2037
bkl	1099	6705	3739	923	2043
df	115	6705	3810	898	1997
mel	1113	6705	3801	925	1979
nv	6705	6705	3751	913	2041
vasc	142	6705	3725	993	1987
Total	10015	46935	26283	6571	14081

After applying data augmentation, a uniform count of 6,705 augmented images per class was generated. We split the data into training, validation, and test sets, utilizing 26,283 images for training, 6,571 for validation, and 14,081 for testing. These splits ensure a balanced representation of each class. This detailed distribution of images allows the model to be well-trained and evaluated, with a balanced exposure to each class, thereby reducing the risk of bias and overfitting.

B. Transfer Learning Model

This study used 13 transfer learning models, such as like DenseNet (121, 169, and 201), ResNet (50V2, 101V2, and 152V2), VGG (16 and 19), InceptionV3, MobileNetV2, and NASNet (Mobile and Large). We got the highest accuracy for the DenseNet121 model. The DenseNet121 architecture, as illustrated in Fig. 3, is a robust deep learning model optimized for image classification tasks through its efficient feature extraction capabilities. The model begins with an input layer that processes images of size 224x224 with three colour

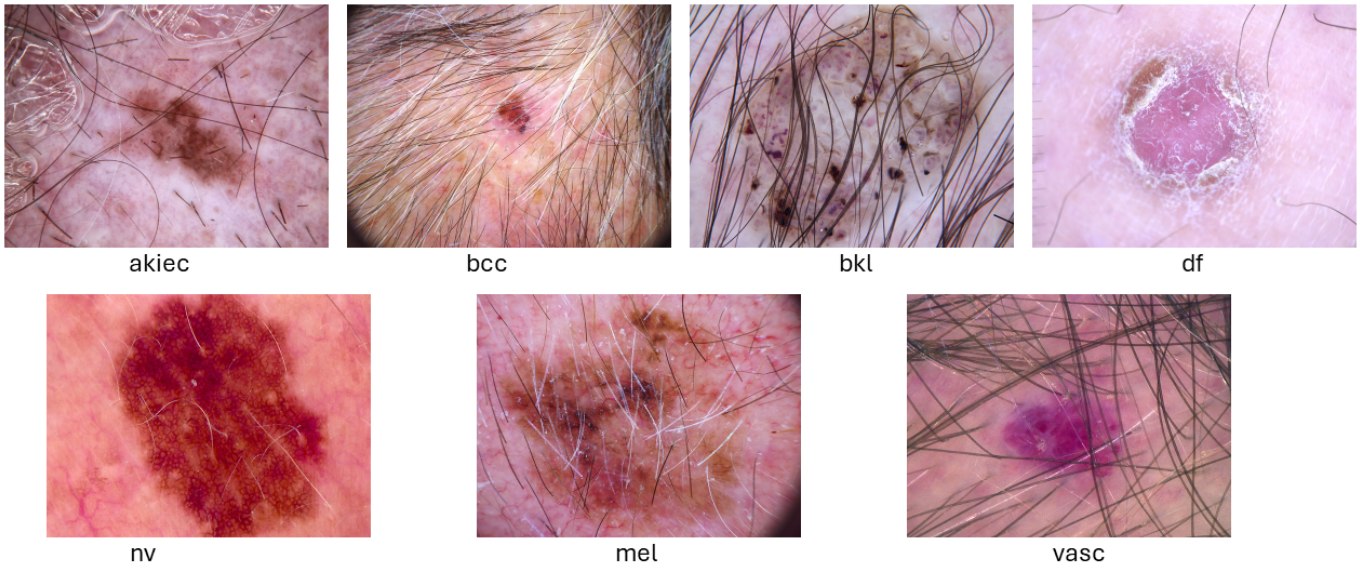


Fig. 2. Seven types of skin cancer.

channels (RGB). The next layer is an initial convolutional layer to capture fundamental features, such as edges. This layer produces a feature map with 64 channels and a spatial dimension of 56x56. The core of the model lies in its dense blocks (D1-D4), where each layer connects to all previous layers within the same block, ensuring that features are reused and information flows efficiently through the network. These dense blocks are interspersed with transition layers (T1-T3), which use 1x1 convolutions and average pooling to reduce the size of the feature maps, thereby maintaining computational efficiency without losing critical information. As the network progresses, the number of channels increases while the spatial dimensions decrease, refining the features extracted. The final layers include a Global Average Pooling (GAP) layer, which condenses the feature map into a vector, followed by fully connected layers that prepare the model for classification. Dropout layers are integrated to prevent overfitting, and the model concludes with a softmax activation function that outputs the probabilities for each class, making DenseNet121 both robust and accurate in image classification tasks.

C. Vision Transformer Model

Fig. 4 compares the architectures of the Vision Transformer (ViT) on the left and DeepViT on the right, highlighting key differences in their approaches to attention mechanisms and processing layers. Both models start with patch embedding and positional embedding to transform input images into a sequence of vectors, enabling the application of transformer-like attention mechanisms. In ViT, the architecture employs a standard multi-head self-attention mechanism followed by normalization and feed-forward layers. This process is repeated multiple times (xN), allowing the model to learn and refine features through self-attention across all patches. The simplicity of ViT lies in its straightforward use of self-attention without any modifications, relying on the depth of the network to capture complex patterns. However, DeepViT presents a modification known as "Re-Attention". After the initial multi-head attention, DeepViT applies normalization and a matrix transformation to refine the attention weights before they are fed into the next layers. This re-attention mechanism aims to stabilize and improve the quality of attention by reinforcing the most critical relationships between patches, potentially leading to better performance in capturing long-range dependencies.

D. Evaluation Metrics

This study used the following metrics to evaluate the performance of the classification model: A True Positive (TP) occurs when a model accurately predicts a positive class, whereas a True Negative (TN) occurs when it correctly predicts a negative class. A False Positive (FP) results when a model erroneously predicts a positive class and a False Negative (FN) happens when a model wrongly predicts a negative class as positive. These metrics were used to calculate the accuracy, precision, recall, and F1-score (see Eq. 1).

$$\begin{aligned} Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\ Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{TP + FN} \\ F1 - score &= \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \end{aligned} \tag{1}$$

Each metric is crucial for evaluating the model's effectiveness in classifying the lesions accurately. Accuracy indicates the overall correctness of the model's predictions. Precision measures the model's ability to correctly identify positive cases, while recall assesses how well the model captures all actual positive cases. The F1-Score, a harmonic mean of precision and recall, provides a balanced measure of the model's performance.

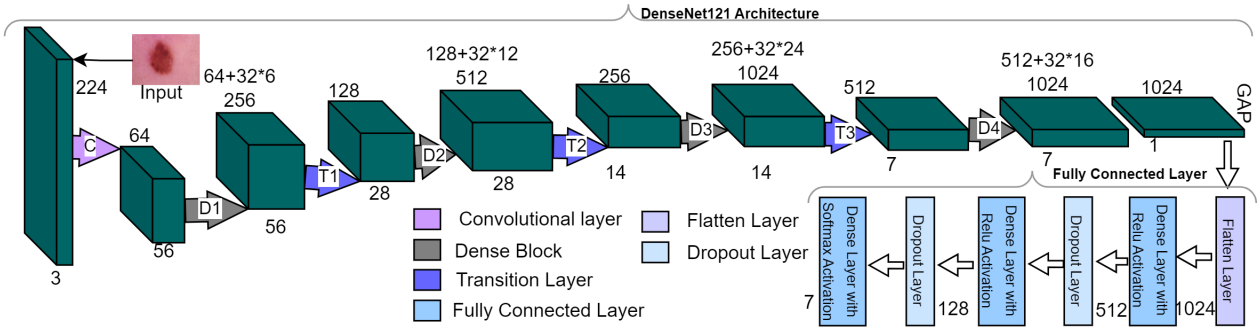


Fig. 3. DenseNet121 architecture.

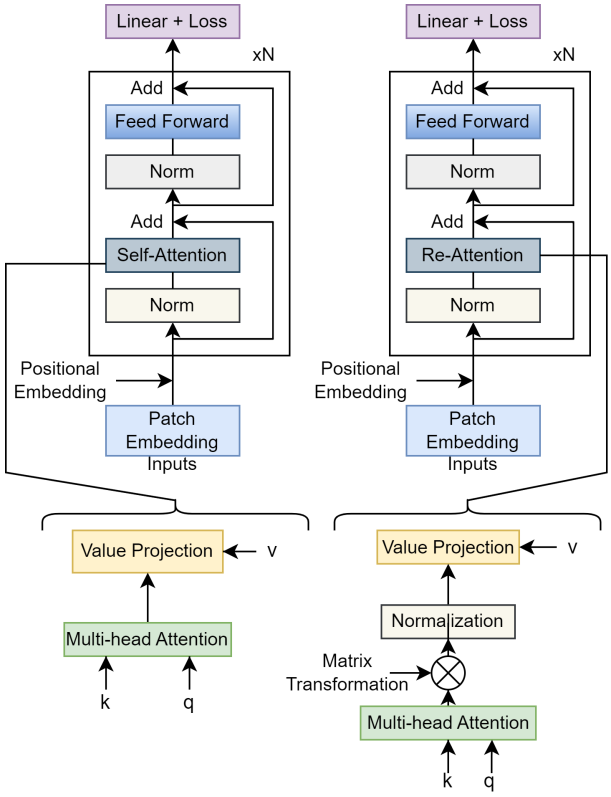


Fig. 4. Comparison between ViT (Left) and DeepViT (Right) model architecture [24].

1) Experimental Setup: The experiment was implemented on the following GPU configurations: NVIDIA GeForce RTX 4090, with 24GB memory. We employed the following parameters to train the transfer learning models: image size (224 x 224), batch size = 32, loss function: "categorical_crossentropy", optimizer: "adam", and two callback functions (ReduceLROnPlateau and EarlyStopping). For the deep vision transformer, we used the following parameters: batch size = 32, epochs = 100, learning rate = 0.00001, gamma = 0.7 and seed = 42, patch size = 32, and embedding dropout = 0.1.

TABLE II. PERFORMANCE COMPARISON OF VARIOUS DEEP LEARNING MODELS1

	Models	Ep	Time (Hour)	Acc	Pre	Re	F1-Score
Transfer Learning	DenseNet121	16	1.94	**0.94**	**0.94**	0.99	**0.94**
	DenseNet169	13	9.87	0.92	0.92	**1.00**	0.92
	DenseNet201	13	10.55	0.92	0.92	**1.00**	0.92
	ResNet50V2	16	3.95	0.92	0.93	0.98	0.93
	ResNet101V2	18	5.46	0.93	0.93	0.99	0.93
	ResNet152V2	26	11.20	0.93	**0.94**	0.99	0.93
	VGG16	12	5.47	0.89	0.91	0.97	0.90
	VGG19	12	5.2	0.87	0.88	0.97	0.87
	InceptionResNetV2	15	3.19	0.91	0.92	0.99	0.91
	InceptionV3	40	7.44	0.90	0.91	0.99	0.90
	MobileNetV2	9	1.49	0.86	0.87	0.98	0.86
	NASNetMobile	30	8.68	0.91	0.91	0.99	0.91
	NASNetLarge	22	8.01	0.90	0.90	0.97	0.90
TF	ViT	80	8.92	0.91	0.92	0.91	0.92
	DeepViT	93	10.56	**0.92**	**0.93**	**0.92**	**0.93**

Ep = Epochs, Acc = Accuracy, Pre = Precision, Re = Recall

IV. RESULT

Table II compares the performance of various deep-learning models employed to classify skin cancers. The models report key performance metrics like accuracy (Acc), precision (Pre), recall (Re), and F1-score, along with the number of epochs (Ep) and training time in hours. For transfer learning, DenseNet121 emerges with the highest accuracy, precision, and F1-score of 0.94 with only 16 epochs and a relatively short training time of 1.94 hours. Additional variations of DenseNet, such as DenseNet169 and DenseNet201, exhibit strong performance while needing longer training durations. ResNet models exhibit impressive performance, especially ResNet152V2, which achieves an optimal trade-off between accuracy (0.93) and precision (0.94) but with the most extended training duration of 11.20 hours over 26 epochs. However, VGG models demonstrate the least accuracy, with VGG16 and VGG19 achieving accuracies of 0.89 and 0.87, respectively. The InceptionV3 and NASNet models demonstrate satisfactory performance, with NASNetLarge and InceptionV3 achieving an accuracy of 0.90. For the vision transformer, the ViT model, while requiring a significantly higher number of epochs (80) and more prolonged training time (8.92 hours), achieves an F1-score of 0.92, indicating robust performance. DeepViT slightly outperforms ViT with a higher F1-score of 0.93, albeit with a longer training time of 10.56 hours.

Fig. 5 illustrates the progression of training and validation loss, as well as accuracy measures, during best 15 epochs for a DenseNet121 model. The training loss, shown by the

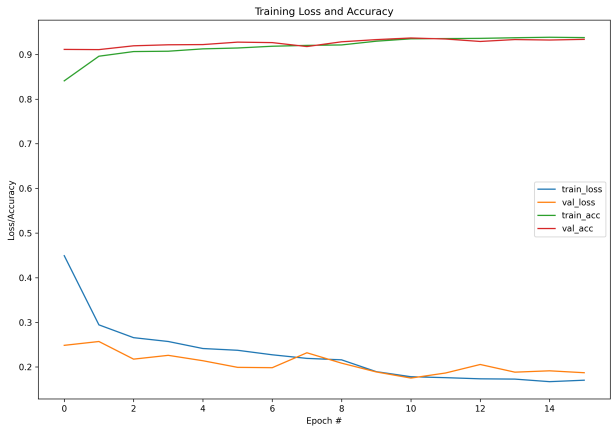


Fig. 5. Training & validation loss and accuracy (DenseNet121).

blue line, initially begins at a very large value and gradually drops, suggesting that the model is progressively learning and improving its performance as time progresses. The validation loss, shown by the orange line, has a similar trend to the training loss but consistently maintains a slightly lower value throughout all epochs. This suggests that the model possesses high generalization capabilities. The training accuracy, shown by the green line, exhibits a rapid initial rise, reaching a plateau by the third epoch, and then maintains a consistently high level for the remaining epochs. The validation accuracy, shown by the red line, has a similar pattern to the training accuracy, indicating that the model consistently performs at a high level on both the training and validation datasets.

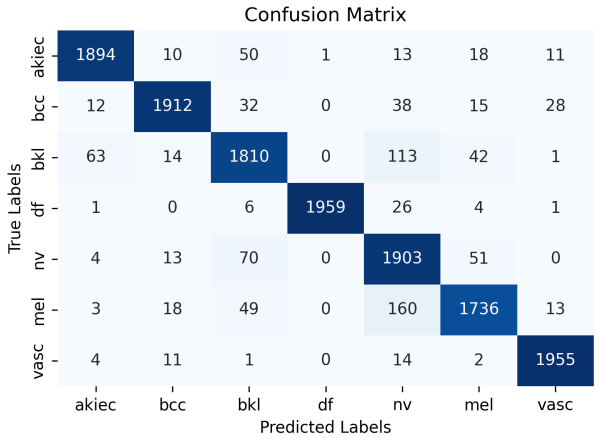


Fig. 6. Confusion matrix (DenseNet121).

Fig. 6 demonstrates the efficacy of a DenseNet121 model on a multi-class skin lesion dataset. The model has high accuracy, accurately categorizing the majority of cases in each category. For example, it accurately recognized 1894 occurrences of akiec, with the majority of misclassifications happening as bkl (50 occurrences). Similarly, the bcc category was accurately identified 1912 times, although some cases of misunderstanding resulted in misclassifications as nv (38 occurrences) and bkl (32 occurrences). The model accurately predicted 1810 cases of bkl. However, it incorrectly categorized some as mel (113) and akiec (63). The df class achieved near-perfect classification, with 1959 accurate predictions and few mistakes. The nv class achieved 1903 accurate classifications, but the model exhibited confusion in distinguishing between mel (70 occurrences) and bcc (51 cases). In the mel class, 1736 occurrences were accurately categorized, whereas 160 were mistakenly classified as nv. Ultimately, the model successfully categorized 1955 occurrences of vasc, with few mistakes. In general, the matrix demonstrates a high level of accuracy in classifying the data; however, there are a few instances where comparable classes were mistakenly identified, namely between mel and nv, as well as between akiec and bkl.

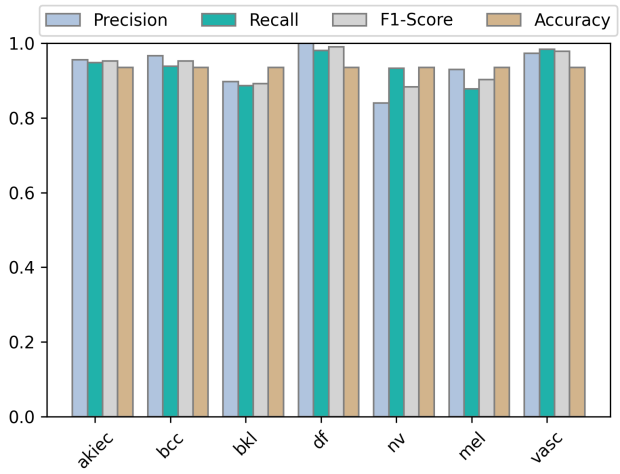


Fig. 7. Precision, recall, F1-Score, accuracy (DenseNet121).

Fig. 7 presents the performance metrics—Precision, Recall, F1-Score, and Accuracy—across different skin lesion classes (akiec, bcc, bkl, df, nv, mel, vasc) for a DenseNet121 model. The figure shows consistently high values across all metrics, demonstrating the model's strong performance across different classes, though slight variations suggest areas for potential improvement in balancing precision and recall for certain classes.

Fig. 8. Training & validation loss and accuracy (DeepViT).

Fig. 8 displays the training and validation loss and accuracy curves over the course of training for a DeepViT model, specifically focusing on the convergence and performance trends across epochs. The training loss (orange line) decreases

steadily, indicating effective learning, while the validation loss (blue line) also decreases but shows more fluctuation, suggesting variability in model performance on unseen data. The training accuracy (red line) improves rapidly and plateaus, reflecting the model's increasing ability to correctly classify training samples. The validation accuracy (green line) follows a similar trend but with smaller fluctuations. This implies that the model generalizes reasonably well but may still be prone to some overfitting, as indicated by the gap between the training and validation losses. Overall, the figure suggests that the model has learned effectively, but the slight divergence between training and validation metrics could indicate room for further tuning to improve generalization.

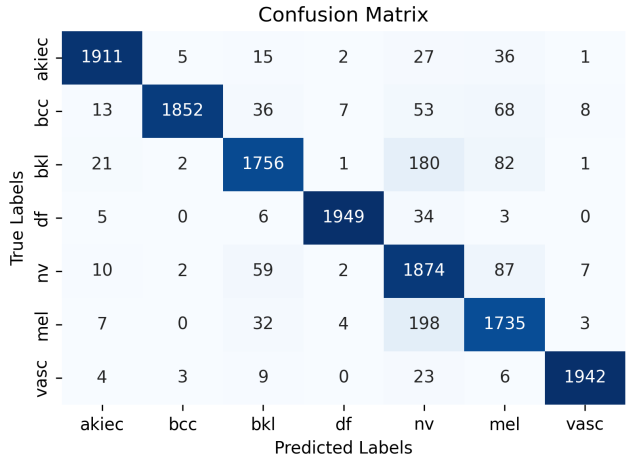


Fig. 9. Confusion matrix (DeepViT).

Fig. 9 illustrates the performance of the DeepViT transformer model in classifying various types of skin cancer. The diagonal elements represent the correctly classified instances for each class, indicating strong model performance, particularly for classes such as akiec, bcc, df, and vasc, with high true positive counts. However, some misclassifications are observed, especially in classes bkl and mel, where the model mistakenly predicts other categories. For example, 180 instances of bkl were misclassified as nv, and 82 instances of mel were also confused with nv.

Fig. 10 presents the performance metrics—Precision, Recall, F1-Score, and Accuracy—of the DeepViT transformer model across different skin cancer classes. The model shows consistently high performance across most classes, with metrics generally above 0.8, indicating effective classification. Notably, the classes akiec, bcc, and vasc exhibit near-perfect scores in all metrics, reflecting the model's strong ability to distinguish these types. These results demonstrate the potential of the DeepViT transformer in enhancing the accuracy of automated skin cancer detection, although further improvements are necessary to address the specific misclassification challenges observed in certain classes.

V. Discussion

Table III presents a comparison of the accuracy achieved by different deep-learning models for skin cancer detection. All the presented articles used the same dataset for multiclass skin

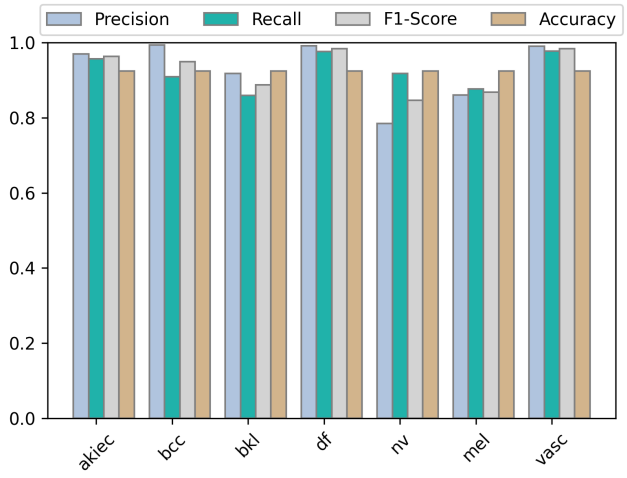


Fig. 10. Precision, recall, F1-Score, accuracy (DeepViT).

cancer classification. It places the DenseNet121 model from the current study in context with other existing methodologies, such as ResNet50, MobileNetV2, DenseNet201, Vision Transformers (ViT), EfficientNet, and others. The reported accuracy ranges from 82.9% to 93.75%. The DenseNet121 model from this study notably achieved an impressive accuracy of 94%, surpassing all the other models listed. This outstanding performance is likely a result of comprehensive data augmentation to tackle class imbalance, efficient transfer learning strategies, and the inherent strengths of the DenseNet121 architecture. These strengths include its deep layers and dense connections, which enable effective feature reuse.

TABLE III. Accuracy Comparison with Prior 7 Types of Skin Cancer Detection Studies

Ref.	Methodology	Accuracy (%)
[11], 2020	ResNet50	90
[12], 2022	MobileNetV2	93.11
[13], 2022	DenseNet201	82.9
[17], 2023	ViT	92.14
[20], 2023	EfficientNetB3	86
[15], 2024	InceptionV3 and DenseNet201	89
[16], 2024	ResNet101	92
[18], 2024	CNN	92.12
[19], 2024	ViT	93.75
[21], 2024	EfficientNet	85
This Study	**DenseNet121**	**94**

VI. Conclusion

This study demonstrates the effectiveness of transfer learning and Vision Transformers (ViTs) in improving the accuracy of skin cancer detection. By addressing the challenges of imbalanced data through extensive data augmentation, we achieved notable results with DenseNet121 and deepViT models, attaining accuracies of 94% and 92%, respectively. These findings underscore the potential of these advanced models in dermatological diagnostics. However, limitations persist, such as the need for large, diverse datasets to generalize across various skin tones and cancer types. Additionally, the computational demands of training ViTs and other deep-learning models remain significant. Future work will focus on overcoming these limitations by exploring hybrid approaches

that integrate the strengths of convolutional neural networks and transformers.

ACKNOWLEDGMENT

The authors acknowledge with thanks the technical support from the Faculty of Computing and Information Technology at the King Abdulaziz University (KAU), Jeddah, Saudi Arabia.

REFERENCES

[1] "Skin cancer," https://my.clevelandclinic.org/health/diseases/15818-skin-cancer, accessed: 20-08-2024.

[2] "Explore cancer statistics," American Cancer Society, https://cancerstatisticscenter.cancer.org/, 2024, accessed: 10-07-2024.

[3] "Cancer stat facts: Melanoma of the skin," National Cancer Institute, https://seer.cancer.gov/statfacts/html/melan.html, accessed: 05-07-2024.

[4] I. Ahmad, F. AlQurashi, E. Abozinadah, and R. Mehmood, "A novel deep learning-based online proctoring system using face recognition, eye blinking, and object detection techniques," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 10, 2021.

[5] I. Ahmad, F. Alqurashi, E. Abozinadah, and R. Mehmood, "Deep journalism and deepjournal v1. 0: a data-driven deep learning approach to discover parameters for transportation," *Sustainability*, vol. 14, no. 9, p. 5711, 2022.

[6] F. Alqurashi and I. Ahmad, "Scientometric analysis and knowledge mapping of cybersecurity." *International Journal of Advanced Computer Science & Applications*, vol. 15, no. 3, 2024.

[7] ——, "A data-driven multi-perspective approach to cybersecurity knowledge discovery through topic modelling," *Alexandria Engineering Journal*, vol. 107, pp. 374–389, 2024.

[8] I. Ahmad, F. AlQurashi, and R. Mehmood, "Machine and deep learning methods with manual and automatic labelling for news classification in bangla language," *arXiv preprint arXiv:2210.10903*, 2022.

[9] ——, "Potrika: Raw and balanced newspaper datasets in the bangla language with eight topics and five attributes," *arXiv preprint arXiv:2210.09389*, 2022.

[10] I. Ahmad and F. Alqurashi, "Early cancer detection using deep learning and medical imaging: A survey," *Critical Reviews in Oncology/Hematology*, p. 104528, 2024.

[11] H. K. Kondaveeti and P. Edupuganti, "Skin cancer classification using transfer learning," in *2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI)*, 2020, pp. 1–4.

[12] P. P. Naik, B. Annappa, and S. Dodia, "An efficient deep transfer learning approach for classification of skin cancer images," in *International Conference on Computer Vision and Image Processing*. Springer, 2022, pp. 524–537.

[13] M. Fraiwan and E. Faouri, "On the automatic detection and classification of skin cancer using deep transfer learning," *Sensors*, vol. 22, no. 13, p. 4963, 2022.

[14] N. Swetha R, V. K. Shrivastava, and K. Parvathi, "Multiclass skin lesion classification using image augmentation technique and transfer learning models," *International Journal of Intelligent Unmanned Systems*, vol. 12, no. 2, pp. 220–228, 2024.

[15] P. Vishnu, S. Krishvadana, C. Rani, R. Khoodeeram, A. Bouridane, and M. Rajesh Kumar, "Melanoma detection with transfer learning," in *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)*, 2024, pp. 1–6.

[16] K. K. Kaveti, M. S. Ravali, V. B. Sravya, and B. Deepthi, "Advancements of skin cancer classification using transfer learning segmentation," in *2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, 2024, pp. 612–616.

[17] M. A. Arshed, S. Mumtaz, M. Ibrahim, S. Ahmed, M. Tahir, and M. Shafi, "Multi-class skin cancer classification using vision transformer networks and convolutional neural network-based pre-trained models," *Information*, vol. 14, no. 7, p. 415, 2023.

[18] T. Tuncer, P. D. Barua, I. Tuncer, S. Dogan, and U. R. Acharya, "A lightweight deep convolutional neural network model for skin cancer image classification," *Applied Soft Computing*, p. 111794, 2024.

[19] G. Yang, S. Luo, and J. Li, "Advancing skin cancer classification across multiple scales with attention-weighted transformers," in *Fourth Symposium on Pattern Recognition and Applications (SPRA 2023)*, vol. 13162. SPIE, 2024, pp. 30–35.

[20] M. Ashfaq and A. Ahmad, "Skin cancer classification with convolutional deep neural networks and vision transformers using transfer learning," in *Advances in Deep Generative Models for Medical Artificial Intelligence*. Springer, 2023, pp. 151–176.

[21] N. Sánchez-Medel, V. Romero-Bautista, R. Díaz Hernández, and L. Altamirano Robles, "Comparison of cnns and vits for the detection of human skin lesions," in *Mexican Conference on Pattern Recognition*. Springer, 2024, pp. 274–283.

[22] P. Tschandl, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," 2018. [Online]. Available: https://doi.org/10.7910/DVN/DBW86T

[23] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti *et al.*, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)," *arXiv preprint arXiv:1902.03368*, 2019.

[24] D. Zhou, B. Kang, X. Jin, L. Yang, X. Lian, Z. Jiang, Q. Hou, and J. Feng, "Deepvit: Towards deeper vision transformer," *arXiv preprint arXiv:2103.11886*, 2021.

Accurate Head Pose Estimation-Based SO(3) and Orientation Tokens for Driver Distraction Detection

Xiong Zhao[1], Sarina Sulaiman[2], Wong Yee Leng*[3]
Faculty of Computing, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia[1,2,3]
Yunnan College of Business Management, Kunming, China[1]

***Abstract*—Driver distraction is an important cause of traffic accidents. By identifying and analyzing the driver's head posture through monitor images, the driver's mental state can be effectively judged, and early warnings or reminders can be given to reduce traffic accidents. We propose a novel dual-branch network named TokenFOE that combines Convolutional Neural Networks (CNN) and Transformer. The CNN branch uses an Multilayer Perceptron (MLP) to infer the image features from the backbone, then generating a rotation matrix based on SO(3) to represent head posture. The Dimension Adaptive Transformer branch uses learnable tokens to represent the head orientation of 9 categories. Integrate the losses of both branches for training, ultimately obtaining accurate head pose estimation results. The training dataset uses 300W-LP, and the quantatitive testing datasets are AFLW-2000 and BIWI. The experiment results show that the Mean Absolute Error is improved by 21.2% and 9.4% compared to the original SOTA model on the two datasets, and the Mean Absolute Error of Vectors is improved by 19.2% and 10.2%, respectively. Based on the model output and calibrated through the camera adapter module, we present the qualitative results on the largest driver distraction detection dataset currently available, the 100-driver dataset, robust and accurate detection results were achieved for four different camera perspectives in two modalities, RGB and Near Infrared. Additionally, the ablation study shows that the model inference speed (21 to 75fps) can be used for real-time detection.**

***Keywords*—Head pose; driver distraction detection; rotation matrix; token; transformer**

I. Introduction

Distracted driving is defined as any behavior in which drivers focus their attention on activities unrelated to driving tasks [1]. The group of drivers who drive for a long time, such as long-distance bus drivers, full-time ride-hailing drivers, Taxi and truck drivers, are more prone to distraction, which is a significant safety hazard. When the driver is distracted, there is usually a phenomenon of abnormal head posture deviating from the direction of vehicle travel. Real time detection can be carried out using the existing monitoring device video in the terminal [2], and timely reminders can be given to the driver when abnormalities occur, which can effectively improve driving safety.

Face Orientation Estimation (FOE) also called Head pose estimation (HPE), is a challenging task in human-computer interaction [3], Human posture detection [4] and attention detection [5], [6]. The FOE problem can be conceptualized as a rotational dynamics challenge involving a 3D rigid body in space. To describe this 3D rotation, various mathematical techniques are available, including Euler Angles, Quaternions, Axis-Angle representations, Rotation Matrices, and Lie algebra. Each of these methods possesses unique strengths and limitations in terms of accuracy, efficiency, and ease of implementation. These advancements are attributed to the utilization of additional data sources such as facial landmark information [7], [8], [9], RGB-depth data [10], multi-task learning approaches [7], [11], and alternative parameterizations for orientation representation [11], [12], [13].

Traditional methods rely on manually extracting facial landmarks, such as the eyes, nose, and mouth, to estimate head pose. These methods are often sensitive to noise and variations in facial features. Deep learning has revolutionized FOE by enabling the direct prediction of head pose from raw images or videos. Deep learning models can learn complex patterns in facial data, leading to more robust and accurate head pose estimation, these methods use extra annotation to help FOE tasks, which usually could get good accuracy. Other methods trying to only use orientation labels, such as HopeNet [14], TriNet [11], FSA-Net [15], and TokenHPE [16], these methods also have much progress but slightly worse than aforementioned methods.

Some methods also try to use transformer architecture, such as TokenPose [17] uses a leaning token to represent a human pose, then regression the whole body key points. Following the same inspiration, TokenHPE [16] tried to use learning tokens to solve the HPE problem, they divided all the orientations into 9 or 11 regions, and every region set a token to predict.

Following these works, we introduce our work, TokenFOE, and try to estimate Head posture by a dual-branch network based SO(3) and Orientation Tokens for driver DDD (Driver Distraction Detection) task. The contributions are summarized as follows:

1) We proposed TokenFOE that adopts the dual-branch network with a joint loss that combines the CNN branch and Transformer branch, we also introduced DAT (Dimension Adaptive Transformer) architecture that could be suitable for any backbone.
2) Extensive experiments on AFLW2000 [18] and BIWI [19] show we achieved new SOTA, especially exceeding other Extra Annotation Free (EAF) methods by a wide margin.
3) To the best of our knowledge, it is the first time that SO(3) and attention mechanism are combined as a dual-branch network in a DDD task.

The paper is organized as follows: Section II introduces some related work in this field, and Section III describes in detail the

model structure of TokenFOE and provides specific important formulas, loss functions, and evaluation metrics. In Section IV, we first introduces the datasets used in this article, and then provides detailed experimental results to demonstrate the effectiveness and robustness of the method. Detailed ablation experimental results are also provided for important parameters. Finally, the limitations and conclusions of this method are discussed in Sections V and VI.

II. Related Works

Prior to the advent of deep learning, estimating facial orientation or pose from RGB images without depth information presented significant challenges. This task involved dealing with a vast representation space encompassing diverse head poses. Over the past decades, numerous methodologies have been proposed to address this challenging problem. In this section, we present a concise overview of the FOE problem.

A. Driver Distracted Detection Approaches

The accuracy of FOE has been greatly improved, and applications based on estimation results have also received more attention and research. Li et al. [20] try to review this research field use electroencephalography. Mou et al. [21] proposed a dual-channel network to try to solve the DDD problem, but it just tested on an early simulated small dataset, and the model's true generalization performance is difficult to verify, with few types of distraction detection. 100-driver [22] is a large-scale, diverse posture-based distracted diver dataset, with more than 470K images taken by 4 cameras observing 100 drivers over 79 hours from 5 vehicles. Gebert P. et al. use an attention module integrated into the network for adaptive feature extraction. Li et al. introduced an efficient system based on a Transformer to detect driving behavior.

B. Extra Information-utilized HPE Approaches

PRNet [23] predicts 2D UV position maps that encode 3D points and utilizes the connectivity of the Basel Face Model (BFM) mesh to construct face models. Research [24] approach combines coarse and fine regression outputs within a deep neural network framework. Meanwhile, SynergyNet [7] explores a synergistic learning process that leverages both 3D Morphable Models (3DMM) and 3D facial landmarks to predict the entire 3D facial geometry, achieving highly accurate results. In contrast, [25] proposes a method that does not require training with head pose labels, instead relying on matching key points between a reconstructed 3D face model and the 2D input image. However, all prediction methods that rely on key points are significantly influenced by the quality of the input image.

Unlike traditional methods that rely on pre-labeled head poses for training, method [25] estimates pose by aligning a reconstructed 3D face model with the 2D input image, bypassing the need for explicit pose labels. However, this approach has a limitation: its accuracy can be affected by variations in image quality, as it depends on precise keypoint detection.

C. Extra Annotation Free HPE Approaches

TokenHPE [16] was a Transformer-based method that is critically aware of minority relationships among facial parts. This approach specifically focuses on learning the intricate relationships between different facial components. To achieve this, they introduce several orientation tokens that are designed to explicitly encode the fundamental orientation regions of the face. Furthermore, they devise a novel token-guided multi-loss function that serves as a guide for the orientation tokens, enabling them to learn the desired regional similarities and relationships. This approach not only enhances the accuracy of head pose estimation but also provides a deeper understanding of the intricate facial geometry. 6dRepNet [13] tackles unconstrained head pose estimation. It overcomes the limitations of prior methods by using a continuous 6D rotation matrix representation for ground truth data. This allows it to learn the full range of head rotations, unlike previous approaches restricted to narrow angles. Additionally, a geodesic distance-based loss function ensures learned rotations adhere to real-world 3D rotation space geometry, boosting accuracy and robustness. LwPosr [26] utilizes a combination of depthwise separable convolutional and transformer encoder layers for efficient and fine-grained head pose prediction.

Overall, these methods have explored various technological routes in the field of FOE and achieved some good results. If new methods want to further improve performance or enhance generalization, they face significant challenges.

III. Method

As shown in Fig. 1, estimating the driver's head posture in the monitored image can help solve DDBR problem.

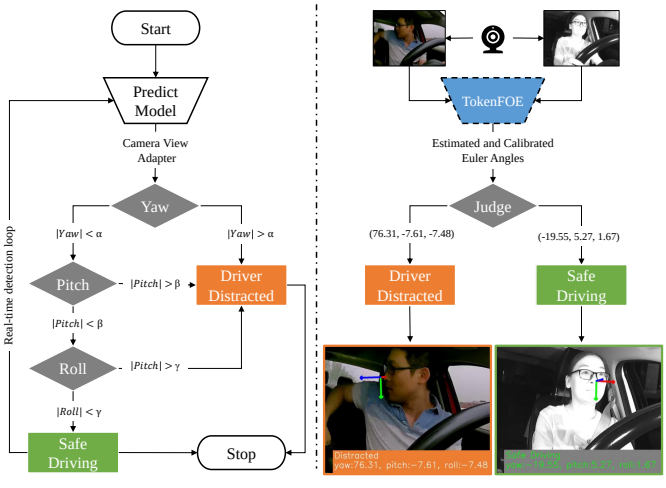


Fig. 1. Left: Overview of proposed driver distraction detection workflow. Right: One RGB and one Near Infrared (NIR) sample of detected results using TokenFOE.

In this section, we first introduce the overview of TokenFOE, then describe every part, including the representation methods of rotation, Gram-Schmidt process, MLP structure, Dimension Adaptive Transformer, and evaluation metric.

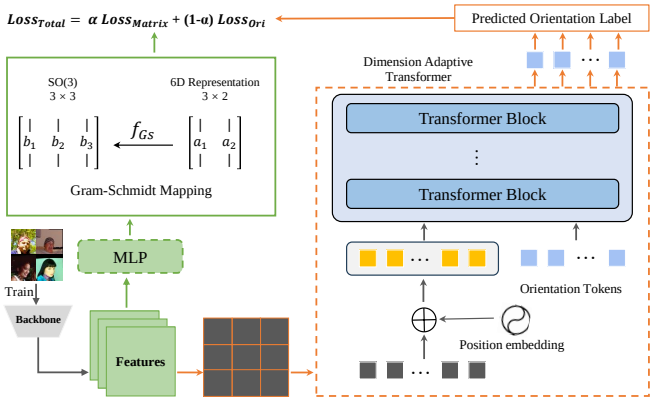


Fig. 2. The pipeline of proposed TokenFOE, a CNN and Transformer fusion dual-branch network based SO(3) and orientation tokens. The main module MLP (*Sec. III-C*), Gram-Schmidt Process (*sec. III-D*) and Dimension Adaptive Transformer (*sec. III-E*) will described respectively.

A. Pipeline of TokenFOE

1) Dual branch architecture: The whole pipeline of TokenFOE is refer to Fig. 2, which first pre-processes the input image (Section IV-B1), then using a pre-trained CNN model to extract image features. The extracted image features are processed in two separate paths. One path is connected to the MLP to predict a 6d representation, follow a Gram-Schimidt process mapped it into a 3×3 rotation matrix belonging SO(3), while the other path adopts the Vit approach to divide features into 14×14 patches, overlay them with position embedding, and send them together with orientation tokens to the transformer block for attention calculation. The calculated results are generated through an fully connected (FC) layer to generate a set of predictions, which are the class of orientation. The calculation results of the two losses are fused using a hyper-parameter α to obtain the final output.

B. Head Orientation and Representation of Rotation

1) Head orientation partitioning: According different orientation euler angles, we could divided head posture to several region. In this work, we followed TokenHPE [16] divided all posture into 9 classes, the illustration refer to Fig. 3.

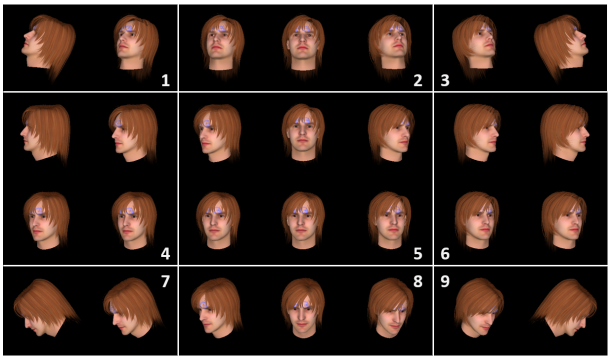


Fig. 3. According the neighbor image similarities, we divided head posture into 9 classes as the schematic diagram.

2) Rotation matrix: There are several methods available to represent a 3D rotation, such as Euler angles, Rotation matrix, Axis-angle, Quaternion, or Lie algebra. Among these, Euler angles are commonly used and intuitive. They decompose a 3D rotation into rotations along the three orthogonal coordinate axes of the object, known as Yaw, Pitch, and Roll. Axis-angle representation and Rotation matrix are closely related and can be converted into each other. If v is a vector in $\mathbb{R}^3$ and k is a unit vector describing an axis of rotation about which v rotates by an angle θ according to the right-hand rule. As we know, an object rotation θ degrees by axis x, y, z could described by followed Eq. 1 and Eq. 2 where the $c_\theta = cos(\theta), s_\theta = sin(\theta), \xi = 1 - cos(\theta)$.

$$R_{(\theta)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_\theta & -s_\theta \\ 0 & s_\theta & c_\theta \end{bmatrix} \begin{bmatrix} c_\theta & 0 & s_\theta \\ 0 & 1 & 0 \\ -s_\theta & 0 & c_\theta \end{bmatrix} \begin{bmatrix} c_\theta & -s_\theta & 0 \\ s_\theta & c_\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$$R = \begin{bmatrix} c_\theta + u_x^2\xi & u_x u_y \xi - u_z s_\theta & u_x u_z \xi + u_y s_\theta \\ u_y u_z \xi + u_z s_\theta & c_\theta + u_y^2 \xi & u_y u_z \xi - u_x s_\theta \\ u_z u_x \xi - u_y s_\theta & u_z u_y \xi + u_x s_\theta & c_\theta + u_z^2 \xi \end{bmatrix} \quad (2)$$

C. MLP Architecture

MLP is a classic artificial neural network model. It consists of an input layer, several hidden layers, and an output layer. In this work, we ultimately chose a single hidden layer MLP module instead of FC according to the ablation study result, with the dimension of the hidden layer set to 768. When using the default RepVgg_b2g4 as the backbone, the complete workflow is shown in Fig. 4.

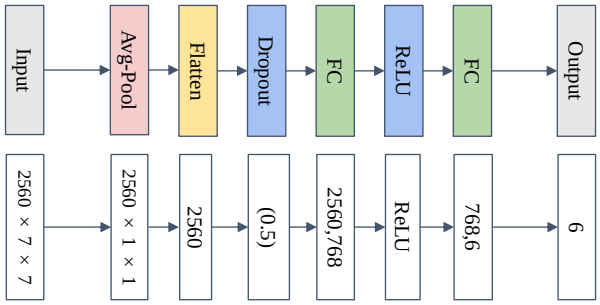


Fig. 4. Upper: MLP architecture with one hidden layer. Lower: the example tensor dimension when use RepVgg_b2g4 as backbone.

D. Gram-Schmidt Process

A set of nonzero vectors $\mu_1, \mu_2, ..., \mu_n$ is called orthogonal if $\mu_i \cdot \mu_j \neq 0$ whenever $i \neq j$. It is orthonormal if it is orthogonal and in addition $\mu_i \cdot \mu_i = 1$ for all $i = 1, 2, 3, ..., n$. We can use the Gram-Schmidt Process to perform orthogonal transformations on the predicted rotation matrix, that is means we could convert the HPE problem to estimate the satisfied rotation matrix.

According TriNet [27] indicated, we could use map function f_{Gs} (refer to Eq. 3) to map the output into $\hat{R}$, furthermore, predict 6 elements could get the best precision and guarantee

the representation space continuity than predict five or nine elements.

$$\hat{R} = f_{Gs}(\begin{bmatrix} | & | \\ a_1 & a_2 \\ | & | \end{bmatrix}) = \begin{bmatrix} | & | & | \\ b_1 & b_2 & b_3 \\ | & | & | \end{bmatrix} \tag{3}$$

The calculate method of the f_{Gs} is using Eq. 4 and Eq. 5.

$$b_i = \begin{bmatrix} N(a_1) & i=1 \\ N(a_2 - (b_1 \cdot a_2)b_1) & i=2 \\ b_1 \times b_2 & i=3 \end{bmatrix}^T \tag{4}$$

$N(\cdot)$ denotes a Normalization function.

$$N(u_i) = \frac{u_i}{\|u_i\|} \tag{5}$$

E. Dimension Adaptive Transformer

ViT [28] is a typical model and also is the first successful vision transformer model. In ViT, an input image undergoes a process of being segmented into smaller patches to avoid the exponential increase in computational complexity caused by too many patches, e.g. the default input image size is $224 \times 224 \times 3$, the author divides the image into $16 \times 16 \times 3$ patches, and maps each patch to a 768 dimensional 1D vector. The total number of patches is $14 \times 14 = 196$.

1) Dimension adaptive layer: In TokenFOE, the object processed by the transformer block is not the original image data, but the features output by the backbones. Different backbones output different feature dimensions and orders, for example, the output of resnet50 is $2048 \times 7 \times 7$, the output of RepVgg_b2g4 is $2560 \times 7 \times 7$, and the output of Swin_base_224 is $7 \times 7 \times 1024$. To be compatible with different feature dimensions without losing information, we introduce a DDT(Dimension Adaptive Transformer), as shown in Fig. 5. The input features undergo dimension adaptation while keeping the number of patches constant, and then are uniformly mapped to 128 dimensions as input for the transformer block. This operation allows our model to match the output of any backbone.

2) Transformer block: Given the 1D token embedding sequence $T = \{[visual], [euler\ angles]\}$ as input, the Transformer encoder learns pose feature representation by stacking M blocks. Each block contains a Multi-head Self-attention (MSA) module and a Multilayer Perceptron (MLP) module. In addition, layer norm (LN) is adopted before every module. Self-attention (SA) can be formulated as Eq. 6.

$$SA(T^{l-1}) = softmax(\frac{T^{l-1}W_Q(T^{l-1}W_K)^T}{\sqrt{d_h}})(T^{l-1}W_V) \tag{6}$$

where $W_Q, W_K, W_V \in R^{d \times d}$ are the learnable parameters of three linear projection layers, T_{l-1} is the output of the $(l-1)_{th}$ layer, d is the dimension of tokens, and $d_h = d \cdot MSA$ is an extension of SA with h self-attention operations which are called "heads". In MSA (Refer to Eq. 7), d_h is typically set to d/h.

$$MSA(T) = [SA_1(T); SA_2(T); ...; SA(h)(T)]W_p \tag{7}$$

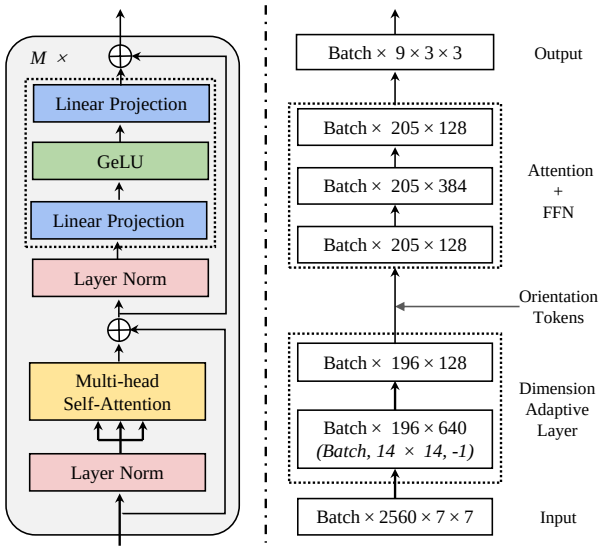


Fig. 5. Left: The architecture of Transformer block that consist of transformer branch. Right: The pipeline and example of dimension adaptive calculation process when use RepVgg_b2g4 as backbone.

Given that spatial relationships are essential for accurate HPE, positional embedding, pos, is added to the visual tokens to reserve spatial relationships, which can be expressed as Eq. 8:

$$[Visual] = \{v_1 + pos, v_2 + pos, ..., v_n + pos\} \tag{8}$$

where, n is the number of patches. Then, we obtain n×1D vectors symbolically presented by [visual] tokens. The position embedding use Eq. 9 followed [29].

$$PE = \begin{cases} \sin(pos/10000^{2i/d_{model}}) & (pos, 2i) \\ \cos(pos/10000^{2i/d_{model}}) & (pos, 2i+1) \end{cases} \tag{9}$$

F. Evaluation Metric

1) Loss function: As we know, for any $R \in SO(3)$ must satisfy $R^T = R^{-1}$, $detR = \pm 1$. Given a specified image x^i from the test data set, there is only one correct rotation matrix converted from the label we called R_{gt}, The model could predict a rotation matrix $\hat{R}$, we want the $\hat{R}$ as closely as possible of R_{gt}, the limitation is $\hat{R} = R_{gt}$, then $\hat{R}R_{gt}^T = I$, I is an identity matrix. we use the Eq. 10 as the loss function followed paper [13].

$$\mathbb{L}_{pose} = \arccos\left(\frac{1}{2}(tr(\hat{R}_i R_{gt_i}^T) - 1)\right) \tag{10}$$

When $R_{gt_i} = \hat{R}_i$, the $\mathbb{L}_{pose}$ equal zero. If the result is not equal to zero, It means our model predicts an incorrect rotation matrix, we could use the result to penalty our model by back propagation algorithm. The final loss function is calculated by Eq. 11, α is a hyper-parameters combine the two branches.

$$Loss_{Total} = \alpha Loss_{matrix} + (1-\alpha)Loss_{ori} \tag{11}$$

2) Evaluate: MAE is a standard metric for HPE, it is defined as Eq. 12.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}(|x_i - \hat{x}_i|) \tag{12}$$

where N is the number of face images and x_i and $\hat{x}_i$ represent the ground truth and predicted pose parameters, respectively.

MAEV is defined as Eq. 13. where N is the number of face images in the dataset and v_i and $\hat{v}_i$ are the vector of ground truth and the predicted result.

$$MAEV = \frac{1}{N}\sum_{i=1}^{N} cos^{-1}\left(\frac{v_i \cdot \hat{v}_i}{|v_i||\hat{v}_i|}\right) \tag{13}$$

IV. Experiments

A. Datasets

We follow the methodologies employed in [13], [15], [11] and utilize well-established public face datasets for training and testing. Specifically, we use the widely recognized 300W-LP [8] as the training set, AFLW-2000 [18] and BIWI [19] for quantitative testing, 100-Driver [22] for DDD task.

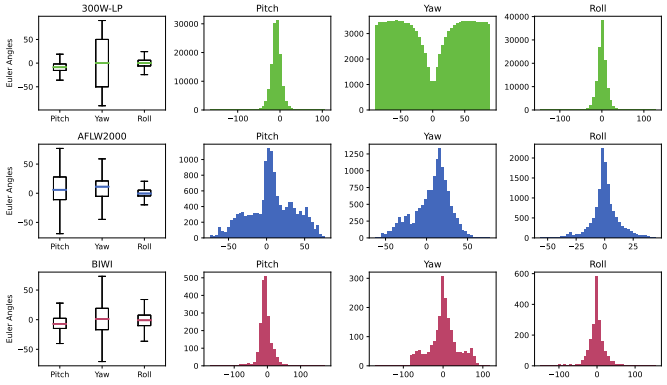


Fig. 6. Box and Hist plots of Training and Testing datasets, statistic by Euler angles. Green(300W-LP [8]), Blue(BIWI [19]) and Red(AFLW2000 [18]).

1) Training: The 300W-LP [8] dataset comprises over 60,000 face samples collected from multiple databases, the statistic results shown in the first line of Fig. 6. We convert ground truth Euler angle into rotation matrices for training.

2) Quantitative testing: AFLW-2000 [18] consists of the first 2000 images from the AFLW dataset. This dataset is annotated with ground truth 3D faces and corresponding 21 landmarks. It offers a diverse range of samples with varying lighting conditions and occlusion levels, providing a robust evaluation environment for our model. BIWI [19] includes more than 15K images of 20 individuals. The head pose range covers about ±75° yaw and ±60° pitch. The statistic data is shown as Fig. 6.

3) Driver distracted detection: 100-Driver [22] is the largest DDD dataset with more than 470K images taken by 4 cameras, observing 100 drivers, over 79 hours from 5 vehicles and including different vehicles, drivers, camera view, and modalities.

B. Model Settings

1) Pre-processing: First, we resize all the images into 300×300 pixels, then random crop to 224×224 pixels as input, we don't use any other augment method. The model uses the pre-trained RepVgg_b2g4 provided by TIMM framework [30] as the backbone, it's one of powerful and popular CNN models. We use Top-Down mode and employ MTCNN [31] to deal with the face location detection task.

2) Hyper-parameters settings: The training and testing environment is Ubuntu 22.04, Python 3.10, cuda 12.6, Pytorch 1.13, and a Nvidia 2080Ti with 11G GPU memory. The core parameters of the training are: the batch size is 64, use the Adam optimizer, the total epochs is 30, and the initial Leaning Rate is set to 1e-4, which is reduced halve when 8^{th}, $16^{th}$and 24^{th} Epoch. All the weights in our model are random initialization and the hyper-parameters α are set to 0.6 by the ablation study result.

C. Experimental Results

Fig. 7 shows that the accuracy on AFLW2000 and BIWI consistently increases with the training progress. There are three periodic low points at 9^{th}, 17^{th} and 27^{th} epoch, respectively, just accompanied by three times decrease in learning rate. The plot also shows an increasing trend until the 30^{th} Epoch, but the growth limit has not been explored yet.

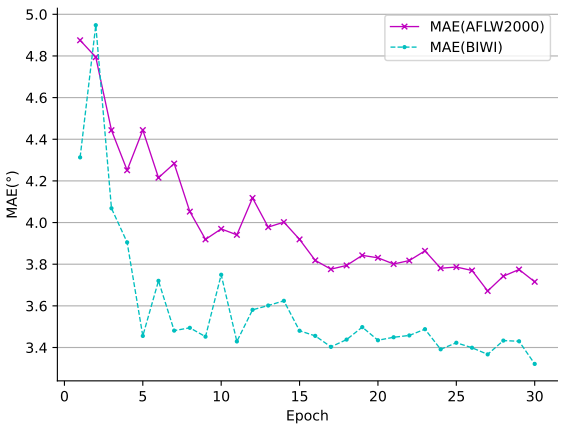


Fig. 7. Per epoch MAE plots tested on AFLW-2000 [18] and BIWI [19].

Follow the old SOTA method, TokenHPE [16], we also show the MAEV scores compared with other popular EAF methods, Refer to Fig. 8, our score exceeds others by a wide margin, specifically, improved by 19.2% on AFLW2000.

HPE is a classical problem that has been studied in many papers. According to literature [16], the methods are divided into two categories: pure HPE and using additional annotated data. Taking SynergyNet [7] as an example, facial key points are used for training. In fact, the coordinate of upper key points already contains head orientation information, which can greatly improve accuracy. As can be seen from Table I, this type of method generally has higher accuracy. Among all methods that do not use additional annotation information, our method has the highest accuracy. We achieved four new SOTA on AFLW-2000 and BIWI. Because TokenHPE [16] is

TABLE I. MEAN ABSOLUTE ERRORS OF EULER ANGLES AND VECTORS ON THE AFLW2000 [18] AND BIWI [19] . ALL METHODS ARE TRAINED ON THE 300W-LP [8] DATASET. COLUMN ‡ INDICATE WHETHER EXTRA ANNOTATIONS FREE. ✓MEANS JUST USING ORIENTATION LABELS FOR SUPERVISING, ✗MEANS USED EXTRA ANNOTATIONS, SUCH AS LANDMARKS. TOKENHPE [16](INDICATED BY UNDERLINE) IS MOST FAMILIAR AND TARGET MODEL. BOLD FONT INDICATE THE SOTA SCORE FOR EVERY COLUMN.

Methods	‡	AFLW2000-Euler				AFLW2000-Vector				BIWI-Euler				BIWI-Vector			
		Yaw	Pitch	Roll	MAE	Left	Down	Front	MAEV	Yaw	Pitch	Roll	MAE	Left	Down	Front	MAEV
EVA-GCN [32]	✗	4.46	5.34	4.11	4.64	-	-	-	-	4.01	4.78	2.98	3.92	-	-	-	-
SynergyNet [7]	✗	3.42	4.09	2.55	3.35	-	-	-	-	-	-	-	-	-	-	-	-
img2Pose [33]	✗	3.43	5.03	3.28	3.91	-	-	-	-	-	-	-	-	-	-	-	-
HopeNet [14]	✓	5.31	7.12	6.13	6.20	7.07	5.98	7.50	6.85	6.01	5.89	3.72	5.2	7.65	6.73	8.68	7.69
FSA-Net [15]	✓	4.96	6.34	4.78	5.36	6.75	6.22	7.35	6.77	4.56	5.21	3.07	4.28	6.03	5.96	7.22	6.40
LwPosr [26]	✓	4.8	6.38	4.88	5.35	-	-	-	-								
Quatnet [34]	✓	3.97	5.62	3.92	4.50	-	-	-	-	4.01	5.49	2.94	4.15	-	-	-	-
Trinet [11]	✓	4.2	5.77	4.04	4.67	5.78	5.67	6.52	5.99	3.05	4.76	4.11	3.97	5.57	5.46	6.57	5.86
TokenHPE-v1 [16]	✓	4.53	5.73	4.29	4.85	6.16	5.21	6.97	6.11								
TokenHPE-v2 [16]	✓	4.36	5.54	4.08	4.66	6.01	5.10	6.82	5.98	3.95	4.51	2.71	3.72	5.41	5.17	6.23	5.60
6DRepNet [13]	✓	3.63	4.91	3.37	3.97	-	-	-	-	**3.24**	4.48	2.68	3.47	-	-	-	-
zhao et al [35]	✓	3.72	4.52	3.16	3.80	-	-	-	-	3.45	4.32	2.75	3.51	-	-	-	-
TokenFOE(Ours)	✓	**3.49**↓ 20.0%	**4.48**↓ 19.1%	**3.04**↓ 25.5%	**3.67**↓ 21.2%	**4.64**↓ 22.8%	**4.37**↓ 14.3%	**5.30**↓ 22.3%	**4.83**↓ 19.2%	3.67↓ 7.1%	**3.91**↓ 13.3%	**2.52**↓ 7.0%	**3.37**↓ 9.4%	**4.79**↓ 11.5%	**4.56**↓ 11.8%	**5.72**↓ 8.2%	**5.03**↓ 10.2%

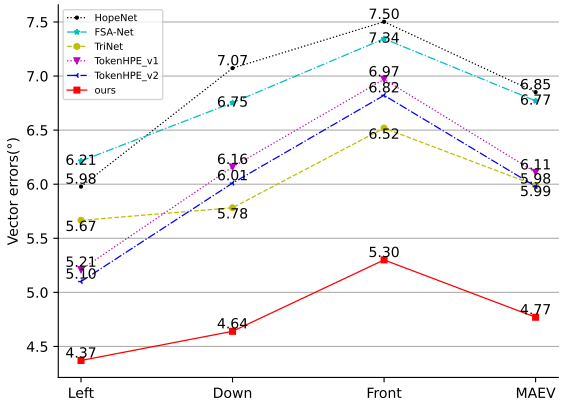


Fig. 8. Comparison with other methods on AFLW-2000 measured by MAEV. Our accuracy exceeds other EAF methods by a wide margin. The lower the better, best viewed in color.

the most familiar with our method, we conducted a careful comparison with it (indicated by underline), using downward arrows(↓) and percentages(%) to indicate the specific values of accuracy improvement.

We conducted extensive ablation experiments to investigate the impact of each module on overall performance.

Table II shows the impact of different backbones on model performance. The results show that the backbone has a significant impact on the final performance. Combining the training and inference speed experimental results in Table III, a trade-off between performance and speed can be made according to specific needs.

TABLE II. ABLATION STUDY OF DIFFERENT BACKBONE. THE OVERALL PERFORMANCE OF THE CNN MODEL IS BETTER THAN THAT OF THE TRANSFORMER MODEL

Backbone	Type	MAE(AFLW2000)	MAE(BIWI)
Swin_base_224	Transformer	5.03	3.44
ResNet50	CNN	4.40	3.69
RepVgg_b2g4	CNN	3.67	3.37

The MLP structure also has a significant impact on accuracy. We compared and tested FC, single hidden layer MLP with different dimensions, and double hidden-layer MLP, and found that the single-layer MLP with dimension 768 had the best performance. All the data refer to Table IV.

TABLE III. ABLATION STUDY RESULTS OF TRAINING AND INFERENCE SPEED. 'M' AND 'MS' ABBREVIATION FOR MINUTES AND MILLISECONDS, RESPECTIVELY

Model	Backbone	Image Size (C,H,W)	Training (M/epoch)	Inference (ms/image)
TokenFOE	Swin_base_224	3,224,224	29	33.9
TokenFOE	ResNet50	3,224,224	13	13.3
TokenFOE	RepVgg_b2g4	3,224,224	31	48.7

TABLE IV. ABLATION STUDY RESULTS OF DIFFERENT MLP ARCHITECTURES. SINGLE HIDDEN LAYER ARE THE BEST MODULE THAN FC AND TWO HIDDEN LAYER ARCHITECTURE

Module	Hidden Layers	Hidden Dimensions	MAE(AFLW2000)
FC	0	-	3.77
MLP	1	128	4.05
MLP	1	256	4.03
MLP	1	512	3.76
MLP	1	768	3.67
MLP	1	1024	3.95
MLP	2	1024,256	4.26

In the end, we tested different loss weights refer to Fig. 9, that is say the different values of α. According to our design, the Transformer branch performs classification tasks, which can improve the overall model accuracy(improved about 11.2%) but is not suitable for independent work. When α is set to 1, means only the CNN branch worked.

So far, our method can accurately estimate the head pose angle, but there is still a problem when facing DDD tasks. The training images are all based on the front profile view as the initial position, and the output Euler angles are also the deviation values relative to this initial position.

However, in the DDD task, the cameras may be arranged in different positions, and due to the influence of the view angle, completely different Euler angles will be output for the same driver posture. This article simply sets up a Camera Adapter layer to solve it. Through cluster analysis for 100-driver [22],

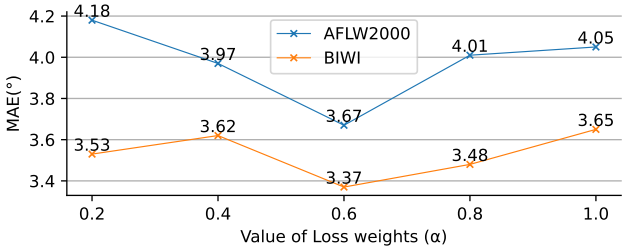


Fig. 9. Ablation study results of model architectures. $Loss_{Matrix}$ is more important than $Loss_{Ori}$.

it was determined that Pitch and Roll directions greater than $\pm 15°$, Yaw direction greater than $50°$, or yaw less than $-75°$ with an absolute value greater than $25°$ are Distracted. Fig. 10 shows the visual detection results of different camera views, modalities, lighting, vehicles, and drivers, further demonstrating the effectiveness and robustness of this method.

V. Limitation and Discussion

In this section, we will discuss the limitations of our method in this article.

First, TokenFOE adopts a dual-branch structure to calculate two sets of Lose separately, which theoretically enhances the model's expressive power. However, compared to CNN, especially lightweight CNN models represented by Mobile-net [36], the computational complexity and resource consumption of the transformer model have significantly increased, which is not cost-effective from an efficiency perspective. The advantage of the transformer scheme is that the model performs better when training with large amounts of data. If we want to deploy the model to a vehicle terminal device for terminal inference, maybe need to fine-tune the model, such as using a lighter backbone and reducing the number of parameters appropriately.

Second, sleeping and yawning is also an important sign of fatigue and distraction, our method can only detect head posture, making it difficult to make quick judgments in such scenarios (refer to Fig. 11).

VI. Conclusion

We proposed a novel dual-branch network named TokenFOE, that combines CNN and Transformer, one is the classic CNN path, and the other is a transformer model based on a self-attention mechanism, the dimension adaptive algorithm suitable uses any pre-trained backbone for the feature extractor.

We train the model on 300W-LP, quantitative test on AFLW-2000 and BIWI. The experiment results show that the MAE score is improved by 21.2% and 9.4% compared to the original SOTA model, and the MAEV score is improved by 19.2% and 10.2%, respectively. Based on the model output and calibrated through the camera adapter module, we present the visualization results on the largest DDD dataset currently available, the 100-driver [22] dataset. Robust and accurate detection results were achieved for four different camera perspectives in daytime (RGB) and Night time(NIR). Additionally, the ablation study shows that the model inference speed (21 to 75fps) can be used for real-time detection.

The main limitation of this method is that the heavy model leads to high training and inference costs, and the computational overhead of the self-attention part is too high. In theory, more tasks can be completed by adding a small number of additional tokens, such as key points. In addition to further improving accuracy and reducing costs, the fusion of multi-task and multi-modality is also a future research direction.

Conflict of Interest

The authors declare no conflict of interest.

Authors' Contributions

Xiong Zhao conceived the main idea, conducted the experiments and wrote the manuscript. Sarina Sulaiman guided the research direction. Wong Yee Leng analyze the results and proposed improvement suggestions. All authors had reviewed and approved the final version.

Funding

This work were supported by Chinese Universities' Industry-University-Research Innovation Fund—BeiChuang Teaching Assistant Project (Phase II) (2021BCF01006), Yunnan Provincial Department to Education Science Research Fund Project (2022J1281; 2024J1352) and The Innovation Team of Intelligent Manufacturing and New Power System Research, Yunnan College of Business Management (2022XKJS02).

References

[1] J. D. Lee, K. L. Young, and M. A. Regan, "Defining driver distraction," *Driver distraction: Theory, effects, and mitigation*, vol. 13, no. 4, pp. 31–40, 2008.

[2] E. Wassef, H. E. Abd El Munim, S. Hammad, and M. Ghoneima, "Robust real-time head pose estimation for 10 watt sbc," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, 2021.

[3] Y. Wang, W. Liang, J. Shen, Y. Jia, and L.-F. Yu, "A deep coarse-to-fine network for head pose estimation from synthetic data," *Pattern Recognition*, vol. 94, pp. 196–206, 2019.

[4] Y. Shu and L. Hu, "A vision-based human posture detection approach for smart home applications," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 10, 2023.

[5] H. Tang, M. Dai, X. Du, J.-L. Hung, and H. Li, "An eeg study on college students' attention levels in a blended computer science class," *Innovations in Education and Teaching International*, pp. 1–13, 2023.

[6] J. Mo, G. Jiang, H. Yuan, Z. Shou, and H. Zhang, "Adaptive target region attention network-based human pose estimation in smart classroom," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 4, 2024.

[7] C. Y. Wu, Q. G. Xu, U. Neumann, and I. C. Soc, "Synergy between 3dmm and 3d landmarks for accurate 3d facial geometry," in *9th International Conference on 3D Vision (3DV)*, ser. International Conference on 3D Vision. LOS ALAMITOS: Ieee Computer Soc, 2021, Conference Proceedings, pp. 453–463. [Online]. Available: ¡Go to ISI¿://WOS:000786496000045

[8] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Z. Li, "Face alignment across large poses: A 3d solution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 146–155.

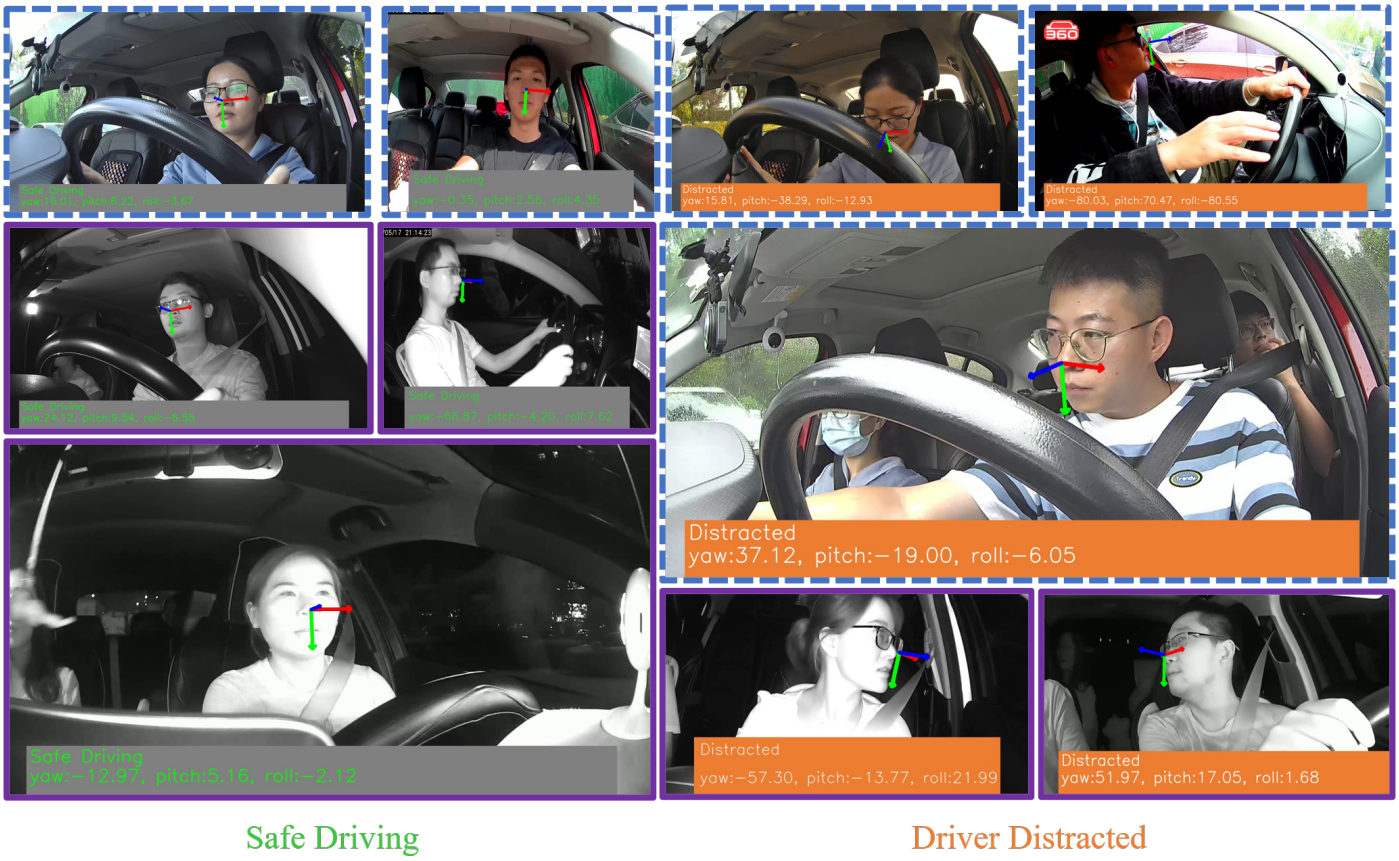


Fig. 10. Visualization of more driver distracted detection results. Blue dashed line, purple bold line, green text and white text indicate Day(RGB), Night(NIR), safe driving and driver distracted, respectively. All images from 100-driver [22].

Fig. 11. Two failure samples.

[9] J. Guo, X. Zhu, Y. Yang, F. Yang, Z. Lei, and S. Z. Li, "Towards fast, accurate and stable 3d dense face alignment," in *European Conference on Computer Vision*. Springer, 2020, Conference Proceedings, pp. 152–168.

[10] Y. Yun, M. H. Changrampadi, and I. Y. Gu, "Head pose classification by multi-class adaboost with fusion of rgb and depth images," in *2014 International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE, 2014, pp. 174–177.

[11] Z. W. Cao, Z. C. Chu, D. F. Liu, Y. J. Chen, and Ieee, "A vector-based representation to enhance head pose estimation," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, ser. IEEE Winter Conference on Applications of Computer Vision. LOS ALAMITOS: Ieee Computer Soc, 2021, Conference Proceedings, pp. 1187–1196. [Online]. Available: ¡Go to ISI¿://WOS:000692171000118

[12] X. Geng and Y. Xia, "Head pose estimation based on multivariate label distribution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, Conference Proceedings, pp. 1837–1842.

[13] T. Hempel, A. A. Abdelrahman, and A. Al-Hamadi, "6d rotation representation for unconstrained head pose estimation," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, Conference Proceedings, pp. 2496–2500.

[14] B. Doosti, S. Naha, M. Mirbagheri, and D. J. Crandall, "Hope-net: A graph-based model for hand-object pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[15] T.-Y. Yang, Y.-T. Chen, Y.-Y. Lin, and Y.-Y. Chuang, "Fsa-net: Learning fine-grained structure aggregation for head pose estimation from a single image," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1087–1096.

[16] C. Zhang, H. Liu, Y. Deng, B. Xie, and Y. Li, "Tokenhpe: Learning orientation tokens for efficient head pose estimation via transformers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, Conference Proceedings, pp. 8897–8906.

[17] Y. Li, S. Zhang, Z. Wang, S. Yang, W. Yang, S.-T. Xia, and E. Zhou, "Tokenpose: Learning keypoint tokens for human pose estimation," in *Proceedings of the IEEE/CVF International conference on computer vision*, 2021, Conference Proceedings, pp. 11 313–11 322.

[18] X. Zhu, Z. Lei, J. Yan, D. Yi, and S. Z. Li, "High-fidelity pose and expression normalization for face recognition in the wild," in

Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 787–796.

[19] G. Fanelli, M. Dantone, J. Gall, A. Fossati, and L. Van Gool, "Random forests for real time 3d face analysis," *International journal of computer vision*, vol. 101, pp. 437–458, 2013.

[20] G. Li, Y. Yuan, D. Ouyang, L. Zhang, B. Yuan, X. Chang, Z. Guo, and G. Guo, "Driver distraction from the eeg perspective: A review," *IEEE Sensors Journal*, 2023.

[21] L. Mou, J. Chang, C. Zhou, Y. Zhao, N. Ma, B. Yin, R. Jain, and W. Gao, "Multimodal driver distraction detection using dual-channel network of cnn and transformer," *Expert Systems with Applications*, vol. 234, p. 121066, 2023.

[22] J. Wang, W. Li, F. Li, J. Zhang, Z. Wu, Z. Zhong, and N. Sebe, "100-driver: a large-scale, diverse dataset for distracted driver classification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 7, pp. 7061–7072, 2023.

[23] Y. Feng, F. Wu, X. Shao, Y. Wang, and X. Zhou, "Joint 3d face reconstruction and dense alignment with position map regression network," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 534–551.

[24] H. Wang, Z. Chen, and Y. Zhou, "Hybrid coarse-fine classification for head pose estimation," *arXiv preprint arXiv:1901.06778*, 2019.

[25] L. Liu, Z. Ke, J. Huo, and J. Chen, "Head pose estimation through keypoints matching between reconstructed 3d face model and 2d image," *Sensors*, vol. 21, no. 5, p. 1841, 2021.

[26] N. Dhingra, "Lwposr: Lightweight efficient fine grained head pose estimation," in *Proceedings of the IEEE/CVF Winter Conference on applications of computer vision*, 2022, Conference Proceedings, pp. 1495–1505.

[27] Y. Zhou, C. Barnes, J. Lu, J. Yang, and H. Li, "On the continuity of rotation representations in neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5745–5753.

[28] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, and S. Gelly, "Vit: An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[29] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, l. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[30] R. Wightman, "Pytorch image models," https://github.com/rwightman/pytorch-image-models, 2019.

[31] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE signal processing letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[32] M. Xin, S. Mo, and Y. Lin, "Eva-gcn: Head pose estimation based on graph convolutional networks," in *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 2021, Conference Proceedings, pp. 1462–1471.

[33] V. Albiero, X. Chen, X. Yin, G. Pang, and T. Hassner, "img2pose: Face alignment and detection via 6dof, face pose estimation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, Conference Proceedings, pp. 7617–7627.

[34] H.-W. Hsu, T.-Y. Wu, S. Wan, W. H. Wong, and C.-Y. Lee, "Quatnet: Quaternion-based head pose estimation with multiregression loss," *IEEE Transactions on Multimedia*, vol. 21, no. 4, pp. 1035–1046, 2019.

[35] X. Zhao, S. Sulaiman, L. Chen, M. Dong, Y. Duo, and H. Song, "Continuity rotation representation for head pose estimation without keypoints," in *Proceedings of the 2023 9th International Conference on Computing and Artificial Intelligence*, 2023, Conference Proceedings, pp. 358–363.

[36] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

Predicting the Most Suitable Delivery Method for Pregnant Women by Using the KGC Ensemble Algorithm in Machine Learning

Pusarla Sindhu, Parasana Sankara Rao
Department of Computer Science and Engineering
GITAM (Deemed to be)University, Visakhapatnam-530045, India

***Abstract*—Maternal and neonatal mortality rates pose a significant challenge in healthcare systems worldwide. Predicting the childbirth approach is essential for safeguarding the mother's and child's well-being. Currently, it is dependent on the judgment of the attending obstetrician. However, selecting the incorrect delivery method can cause serious health complications both in mother and child over short-time and long-time. This research harnesses machine learning algorithms' capability to automate the delivery method prediction process. This research studied two different stackings implemented in machine learning, leveraging a dataset of 6157 electronic health records and a minimal feature set. Stack1 consisted of k-nearest neighbors, decision trees, random forest, and support vector machine methods, yielding an F1-score of 95.67%. Stack 2 consisted of Gradient Boosting, k-nearest neighbors, and CatBoost methods, which yielded 98.84%. This highlights the superior effectiveness of its integrated methodologies. This research enables obstetricians to ascertain the delivery method promptly and initiate essential measures to ensure the mother's and baby's safety and well-being.**

Keywords*—Delivery method; stacking; neonatal mortality; KGC ensemble algorithm

I. Introduction

Maternal and neonatal mortality can be significantly reduced by carefully selecting the most appropriate delivery method based on the mother's health, pregnancy complications, and fetal conditions [1]. Mode of childbirth has become a significant issue for obstetricians, health authorities, and mothers. Over the current years, there has been a global rise in cesarean delivery rates, contrary to the World Health Organization's guidelines advocating for rates below 15% [2]. However, determining an appropriate cesarean delivery rate and the impact it has on maternal and infant well-being remains a topic of debate within the obstetrics community [3]. Progressive medicalization can be attributed partly to the ever-rising rates of cesarean sections [4].

Medical technology advancements such as elective or planned Caesarean, emergency cesarean section, forceps delivery, and vacuum extraction have resulted in a safer experience during childbirth. In a cesarean delivery, also known as a C-section, surgeons perform a laparotomy (an abdominal incision in the mother) followed by a hysterotomy (an incision in the uterus) to facilitate the birth of the baby [5]. Broadly, experts classify cesarean sections into three categories: elective or planned, emergency, and cesarean on demand [6].

Obstetricians opt for cesarean delivery or C-section if the fetus inside the mother's womb is in an unusual position, has very little amniotic liquid, possessing multiple fetuses or many other reasons. Choosing the wrong delivery technique may impose immediate and long-term health complications on both mother and child. Seeking the help of technology can help obstetricians make decisions accurately. Machine learning can guide the obstetrician in predicting the probable chances of the type of delivery, which can educate the mother on a safer mode of birth.

The elective cesarean is a prepared delivery when maternal or fetal indications arise in the antepartum period opted before the ongoing labor. On the other hand, medical professionals perform an emergency cesarean when they observe symptoms that emerge during labor, requiring immediate medical intervention. Healthcare providers perform a cesarean on demand when women specifically request it, as indicated by its name. Cesarean delivery is performed for patients in conditions like when the mother possesses more than one fetus when the baby is in a breech position, when the mother has severe health conditions like diabetes or pre-eclampsia, when labor does not progress any further, when elderly primigravida or dystocia or maternal HIV occurs, C-sections can be a life savior. Assisted deliveries like forceps and vacuum extraction can provide extra support during childbirth for the protection of both mother and baby. Forceps provide traction on the baby's head and vacuum extraction using suction to assist delivery.

A well-chosen birthing technique-whether vaginal birth, cesarean section, or other interventions can help avoid potential risks, ensuring safer childbirth outcomes. The gynecologist chooses the delivery method by evaluating various obstetric characteristics like the number of fetuses and the medical history of the pregnant woman, such as diabetes and blood pressure. Personalized care, guided by healthcare professionals, ensures that the delivery method aligns with maternal and fetal needs, reducing the likelihood of complications and improving overall survival rates for mothers and newborns.The gynecologist selects the delivery method by evaluating various biological factors of the mother, such as age, medical history, and other relevant health conditions.

In this research, the proposed algorithm is an ensemble of K-Nearest Neighbors (KNN), CatBoost (CB), and Gradient Boosting (GB). The algorithms have been wisely chosen for their complementary strengths. KNN is helpful for less noisy datasets because of its simplicity and interoperability,

while CatBoost handles categorical data efficiently, even on imbalanced datasets.GB excels in capturing complex relationships and delivering high predictive accuracy. The stack of these classifiers offers flexibility to explore a diverse learning approach to optimize model performance with hyperparameter tuning.

This research focuses on tackling the existing literature gap by developing a predictive model that integrates multiple clinical features to define the mode of childbirth. The approach utilizes a stacked ensemble of ML algorithms to optimize performance through hyperparameter tuning, providing more accurate predictions. The novelty of this study lies in diverse classifiers, which significantly enhance the model's performance metrics.

The main contributions of this research include

- A detailed literature survey on the type of algorithms used in the recent studies.
- Model can handle large datasets with more features accurately.
- The study employs three classifier algorithms stacked to ensure more precise and reliable decisions.

The paper is structured as follows: Section II provides a review of related literature, Section III presents the methods and materials used, Section IV describes the experimental process Section V analyzes the results obtained, and finally, Section VI concludes the study and suggests potential future research directions.

II. Literature Survey

Recent studies have witnessed the growing importance of machine learning by various authors for pregnancy outcomes and optimizing medical decisions.Fernández et al. [3] investigated algorithms like support vector machines(SVM), Random forest(RF), and the Multilayer Perceptron (MLP) to forecast the delivery type among three classes, namely C-section, euthocic vaginal, and instrumental deliveries. HGSORF, which applies the Henry Gas Solubility Optimization algorithm with Random Forest was designed to predict C-sections, demonstrating the potential of decision-making [7]. Khan et al. [8] investigated how machine learning techniques transform gynecological healthcare, aiming to enhance diagnostics and the challenges faced. Lestari et al. [9] conducted a comprehensive review on anticipating pregnancy-related complications. Tiruneh et al. [10] organized a broad review to compare Pre-Eclampsia Prediction using regression models and machine learning algorithms.

Islam et al. [11] performed a systematic review on ML uses to predict pregnancy outcomes, identifying gaps in the existing approaches and proposing a research agenda for future work.Kolasa et al. [12] have done a diverse review of the usage of ML algorithms in health care. Mas-Cabo et al. [13] studied algorithms like Multilayer Perceptron(MLP) and Artificial Neural Network(ANN) to forecast anticipated labor in women with early labor symptoms through the analysis of electrohysterogram (EHG) signals of the uterus. ANN was employed to estimate the success of labor induction, analyzing the uterine EHG signals [14]. The authors achieved different objectives with ML, such as classifying placenta cells [15], developing pregnancy disorder in the first-trimester prediction [16], and evaluating cesarean delivery risk in term nulliparous.

XGBoost (Extreme Gradient Boosting) has been extensively used to solve predictive modeling tasks by constructing a series of decision trees to rectify previous errors and avoid overfitting. To illustrate, Sultan [17] experimented with different algorithms to identify the most suited algorithm to classify cesarean section deliveries. Xi [18] worked towards predicting the large gestational age neonates in parturients exposed to radiation using machine learning.Yu [19] evaluated using machine learning algorithms for preterm birth forecast in singleton pregnancies through time-series data [20].

Stacking has been popular machine learning, widely used in various applications like the detection of thyroid diseases [21], software bug prediction [22] predicting childbirth approach, preterm birth prediction [23] used for various applications. Islam et al. [4] focused on the features suited best for the prediction of the delivery technique using algorithms like RF, SVM, Decision Tree (DT), K-Nearest Neighbors (KNN), and stacking classifier (SC). Yang and Shami [24] explored hyperparameter optimization across most machine learning algorithms, emphasizing its critical role in enhancing predictive performance.

III. Methods and Materials

In this study, we assessed the appropriateness of utilizing various ML algorithms for anticipating the mode of childbirth across six classifications: CES Programmed(Elective Cesarean), Emergency Cesarean, Eutocic delivery, Forceps delivery, vacuum extraction, and Epistiomy.

The dataset comprised instances of women presented for childbirth at four public hospitals across three distinct autonomous regions in Spain in 2014 [25]. One hundred and sixty-one personal health and medical features were recorded from each mother and the fetus; few details were gathered in antepartum and remaining in intrapartum.

A. Dataset Description

The Target variable Type of birth was noted as the significant result of the labor. The Target Variable has been classified as one of the classes CES Programmed, Elective Cesarean, Eutocic delivery, Forceps delivery, vacuum extraction, and Epistiomy. Medical circumstances surrounding the mother and fetus were given utmost priority in deciding the type of delivery technique for their well-being.

B. Data Preprocessing

This process is crucial for ensuring the accuracy and effectiveness of the models built using the data and preprocessing, including data reduction, preparation, and balancing techniques. The dataset contained 6157 data records with 161 attributes. Data reduction is a crucial preprocessing approach that can decrease the size and complexity of a dataset while preserving its essential information. It involves removing or merging redundant or irrelevant features, removing noise and outliers, and transforming the data into a more convenient representation.

Data reduction aims to make the dataset more manageable and accessible without sacrificing important patterns or relationships. Dimensionality reduction helps machine learning algorithms function more effectively and predict better. We discarded features with minimal or no impact (e.g. OLIGOAMNIOS, Isoimmunization, MIOMECTOMY, and many more). Other aspects, such as the sex of the fetus and fetal admission to the ICU, are only identifiable after the delivery process. Islam and his colleagues conducted structured interviews with 111 features to obtain information on their relevance [4]. The weighted average score, which ranges from 1 to 5, was calculated across 111 features by averaging each score and considering their assigned weight. The importance or significance of the feature can determine the weight. Thirty-two features scored above 1.5 on average, while 79 scored below 1.5. Setting a threshold of 1.5, the initial set comprised 32 features, with the remaining 79 considered less significant. Table I contains the 32 attributes of the dataset selected after preprocessing.

Data reduction was achieved by removing 18 duplicate records from the dataset. We replaced missing values for the numerical features with their mean, and the mode was substituted for categorical features. Data preparation involves initializing input data to make it suitable for effective analysis by algorithms. The MinMaxScaler transforms the dataset's features, allowing each feature to have an equal impact on the predictive capability by scaling to an interval of 0 to 1. Subsequently, the normalize function applies L2 normalization, which modifies the feature vectors to have a unit norm, improving the model's capacity to handle different magnitudes across the data.

The dataset suffered from a strong inequality between the superior class Eutocic and the minority class Epistiomy. To mitigate overfitting and prevent the model from favoring the majority class excessively in terms of accuracy and frequency over the minority class, random oversampling was employed to balance the class distribution until achieving a 1:1 ratio. Random oversampling duplicates the samples from the minority class and introduces them to the superior class to achieve a balanced representation of majority and minority samples in the dataset.

C. Algorithms

1) K-Nearest Neighbors: This non-parametric algorithm assigns labels or predicts values depending on the majority class of its closest neighbors in the feature domain.

2) Gradient Boosting: This machine learning ensemble technique constructs predictive models by merging weak learners, which are decision trees, into a strong ensemble model.

3) CatBoost: Categorical boosting, also known as CatBoost, stands out for its efficient handling of categorical features. This algorithm automatically handles categorical variables without extensive preprocessing, making it convenient for real-world datasets with common categorical features.

IV. Methodology

The algorithms were created, trained, and evaluated using Jupyter Notebook, a scientific program development and open-source development platform implemented in Python utilizing the Scikit-learn package. From the data, 80% is randomly utilized for the trained model, and the remaining 20% is employed for testing and evaluating each model.

A. Experimental Setup

Each algorithm is tuned for the hyperparameters using Grid search with five-fold cross-validation and accuracy as the performance metric for tuning the hyperparameters. Through an exploratory research approach, the hyperparameters have been selected after testing over a broad range of configurations. This gives the flexibility in identifying optimal parameters based on the dataset's performance rather than predetermined theoretical values.

1) K-Nearest Neighbors: Several configurations of the KNN algorithm were used as it doesn't build the model explicitly during training. A general rule doesn't exist that predicts the optimal value of the parameters, which in turn is dependent on the dataset characteristics and must be found empirically. A grid search is employed as a tuning technique to identify the optimal values of the hyperparameters. Grid Search has been applied to the K-Nearest Neighbors (KNN) algorithm, tuning the hyperparameters across the following ranges: the number of neighbors (n_neighbors) from 1 to 20, the weighting metrics ("uniform" and "distance"), and the distance metrics ("Euclidean", "Manhattan", "Minkowski").

- n_neighbors: This parameter dictates the number of neighboring data points examined during prediction, influencing the model's decision boundary flexibility. A smaller "n_neighbors" value can make the model more susceptible to noise, potentially causing overfitting. Conversely, a larger "n_neighbors" value can make the model too generalized, possibly overlooking local patterns in the data.
- metric: This parameter defines the distance metric used to measure the distance between points in the feature space. Various distance metrics lead to varied notions of proximity between points, which can affect the performance of the K-nearest neighbors algorithm.
- distance: The weights parameter steers the decision of data point classification in KNeighborsClassifier. Exploring various weight configurations can boost the model's effectiveness, mainly when working with imbalanced datasets or fluctuating feature significance.

The best parameters the classifier finds with grid search for this algorithm are {"n_neighbors": 5, "metric": "manhattan", "weights": "uniform"}.

2) Gradient Boosting: Grid Search has been applied to the algorithm, tuning the hyperparameters across the following ranges considering the number of estimators from 10 to 200, learning rate from 0.01 to 1.0, and maximum depth of trees from 1 to 20.

- n_estimators (Number of Estimators): This parameter dictates the number of boosting trees built in the ensemble. A higher number of trees generally leads to a more expressive model, potentially capturing intricate patterns in the data.

TABLE I. FEATURES OF THE DATASET

Feature	Description	Type
AGE	Age of the individual	Numerical
ALCOHOL	Alcohol consumption during pregnancy	Categorical
AMNIOCENTESIS	A medical procedure involving the extraction of amniotic fluid for various diagnostic purposes, such as genetic testing.	Categorical
AMNIOTIC LIQUID	Characteristics of amniotic fluid	Categorical
ANESTHESIA	Administration of anesthesia during childbirth or related medical procedures.	Categorical
ART	Presence of assisted reproductive technology in the conception of the pregnancy.	Categorical
ART MODE	Methods to Assisted Reproductive Technology	Categorical
BMI	Body Mass Index of the individual	Numerical
CARDIOTOCOGRAPHY	Monitoring the fetal heartbeat and uterine contractions during labor.	Categorical
COMORBIDITY	Presence of one or more additional disorders or diseases alongside the pregnancy	Categorical
COMPLICATIONS	Medical issues or difficulties during pregnancy or childbirth.	Categorical
EPISIOTOMY	A surgical incision made during childbirth to widen the opening of the vagina.	Categorical
FETAL INTRAPARTUM PH	Measurement of the acidity or alkalinity of the fetal blood during labor.	Categorical
GESTATIONAL AGE	The age of the fetus in weeks, calculated from the beginning to end of the menstrual period.	Numerical
HEIGHT	Height of the individual	Numerical
INDUCTION	Method of labor induction	Categorical
KG INCREASED PREGNANCY	Increase in weight during pregnancy	Numerical
MATERNAL EDUCATION	Level of education of the mother	Categorical
MISCARRIAGES	Number of miscarriages	Numerical
NUMBER OF PREV CESAREAN	Number of previous cesarean deliveries	Numerical
OBSTETRIC RISK	Risk factors associated with pregnancy	Categorical
OXYTOCIN	Use of oxytocin during labor	Categorical
PARITY	The count of a woman giving birth to a fetus crossing the gestational age of 24 weeks or more.	Numerical
PREINDUCTION	Usage of medical interventions to initiate labor before it starts spontaneously.	Categorical
PREVIOUS CESAREAN	Indication of whether the mother had a cesarean section in a previous pregnancy.	Categorical
PREVIOUS PRETERM PREGNANCIES	Number of previous preterm pregnancies	Numerical
PREVIOUS TERM PREGNANCIES	Number of previous full-term pregnancies	Numerical
ROBSON GROUP	Robson classification group	Categorical
SMOKING	Smoking status during pregnancy	Categorical
START ANTENATAL CARE	Timing of initiation of medical care and attention during pregnancy.	Categorical
SUBSTANCE ABUSE	History of harmful substances (e.g., drugs or alcohol) during pregnancy.	Categorical
WEIGHT	Weight of the individual	Numerical

- learning_rate (Learning Rate): The learning rate controls the contribution of each tree to the final prediction. A higher learning rate speeds convergence but risks overshooting, while a lower rate requires more iterations but enhances generalization.
- max_depth (Maximum Depth of Trees): This parameter determines the maximum depth allowed for each tree in the ensemble. Deeper trees can capture more detailed data features but may also result in overfitting as they learn the noise in the training data.

The hyperparameters by the grid search for this algorithm are "n_estimators": 90, "learning_rate": 0.5, "max_depth": 10.

3) CatBoost: A systematic exploration of different parameter combinations was conducted to optimize the training of CatBoost models involving varying key hyperparameter ranges, namely, the number of estimators ("n_estimators") from 10 to 200, the depth ("depth") from 1 to 16, and the learning rate ("learning_rate") from 0.01 to 1.0.

- n_estimators (Number of Estimators): This defines the number of trees (boosting iterations) to be built during training. Increasing the number of estimators can lead to a more complex model, potentially improving performance, but it may also increase training time and the risk of overfitting
- learning_rate (Learning Rate): The learning rate controls the step size at each iteration during the gradient descent optimization process. It adjusts the model weights in response to the error gradient.
- max_depth (Maximum Depth of Trees): This parameter sets the maximum depth allowed for each tree in the ensemble. Deeper trees can capture more complex relationships in the data but may also lead to overfitting.

The hyperparameters by the grid search for this algorithm are "n_estimators": 100, "learning_rate": 0.5, "max_depth": 10.

4) Stacking: An ensemble learning method combines to generate a new training set for a meta-classifier based on the predictions of multiple classifiers. Each classifier is trained on the entire training set individually, and the meta-classifier is learned from the predictions made by the base models. Fig. 1 illustrates the stacking classifier's architecture. The initial training data (X) had 6157 samples and 2 features. Three M different models (M = 3) are trained on X, and their predictions (y) are combined to generate a data set X^2 for the level 2 model. A strong SC was proposed in which KNN, GB, and CB with the hyperparameters were the base classifiers.

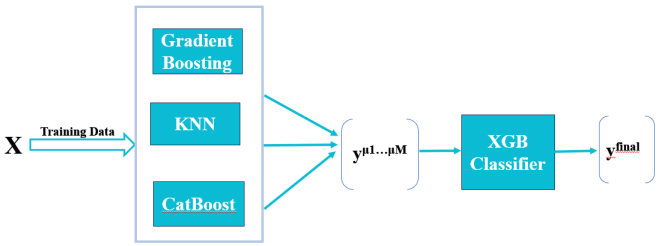


Fig. 1. Proposed KGC ensemble architecture.

5) Proposed Algorithm: KGC Ensemble Algorithm: A KGC ensemble algorithm has been developed using the stacking of KNN, GB, and CatBoost algorithms with the hyperparameters found in the GridSearch. This algorithm excels beyond the standard ML algorithms with default parameters (Stack 1). The primary goal of the stacking is to minimize data variance and optimize its suitability for machine learning models.

K-Nearest Neighbors (KNN) starts the process by classifying data points based on the proximity to other points in the dataset. In KNN, each data item is assigned to the class frequency among k nearest numbers where k is predetermined.

$$h(x) = \text{majority\_label}\left(\{y_i \mid x_i \in N_k(x)\}\right)$$

where:

- x is the data point to classify,
- $N_k(x)$ represents the k-nearest neighbors of x,
- and *majority_label* denotes the most frequent label among these neighbors.

The algorithm classifies based on the majority class contribution among neighbors, effectively performing similarity-based classification.

Further, the dataset undergoes similar training under Gradient Boosting (GB). GB builds an ensemble of decision trees that sequentially trains every tree to rectify the errors of its predecessor. The process continuously minimizes a loss function, making the model more accurate. Prediction is a weighted sum of the predictions of all trees, and the function h can be expressed as:

$$h(x) = \sum w_i f_i(x)$$

where:

- $f_i(x)$ are the individual decision trees,
- and w_i are the weights assigned to each tree's prediction.

Finally, CatBoost, an advanced boosting algorithm, is applied. CatBoost builds an ensemble of trees, similar to GB, but uses ordered boosting and permutation techniques to ensure unbiased and robust learning. The hypothesis function h in CatBoost can be represented as:

$$h(x) = \sum \alpha_i T_i(x)$$

where:

- $T_i(x)$ are the individual trees,
- and α_i are the weights assigned to each tree's output.

The ensemble of these models KNN, GB, and CatBoost-work to improve the overall performance, reduce variance, and provide a robust solution for classification tasks. The pseudocode of the ensemble is shown in Algorithm 1.

Algorithm 1 KGC Ensemble Algorithm:

Input: Initial training dataset $D_s = \{(x_i, y_i)\}_{i=1}^{k}$
Output: An ensemble classifier H
$B_1, B_2, \ldots, B_L$: Base classifiers
M: Meta classifier
p_{il}: Predicted values output by base classifiers B_i for data sample x_i
Step 1: for $l \leftarrow 1, 2, \ldots, L$ do
Train base classifier B_l on D_s to obtain h_l
Step 2: for each data sample x_i in D_s do
for $l \leftarrow 1, 2, \ldots, L$ do
$p_{il} = h_l(x_i)$ *(Predict using base model h_l)*
Step 3: Form the new dataset D_{new} by augmenting D_s with the predictions p_{il}
$D_{\text{new}} = \{(x_i, p_{i1}, p_{i2}, \ldots, p_{iL}, y_i)\}_{i=1}^{k}$
Step 4: Train the meta classifier M on D_{new}
return $H(x) = M(x, h_1(x), h_2(x), \ldots, h_L(x))$

6) Performance evaluation of algorithms: The following metrics are evaluated at each stage of the experimentation.

- Precision: Precision calculates the number of predictions made for a class that belongs to the class.
- Recall: Recall calculates the number of estimates made for a class amongst all the cases of that class present in the dataset
- F1-Score: The F1-score is the weighted average of precision and recall, typically ranging from 0 to 1.

V. Results and Discussion

Two stack ensembles are constructed to evaluate the performance of the different model configurations. The first stack consisted of the classifiers DT, KNN, RF, and SVM balanced using the sampling technique ADASYN. In contrast, the second stack contained KNN, GB, and Catboost, each tuned with the optimized hyperparameters with the random oversampling technique. The experimental setup involved comparing the predictive performance of both stacks on the same validation dataset, with results being analyzed to assess the overall effectiveness of the ensemble.

Results provided by the proposed stack showcase an excellent predictive capability to classify the type of delivery method correctly. Even though the individual methods performed well, the proposed algorithm combines the pitfalls of each method and showcases good results compared to the individual models. Fig. 2 represents the stack's performance comprising the models SVM, DT, KNN, and RF balanced using the ADASYN balancing technique, and the metrics of the proposed KGC Ensemble balanced using random oversampling are depicted in Fig. 3.

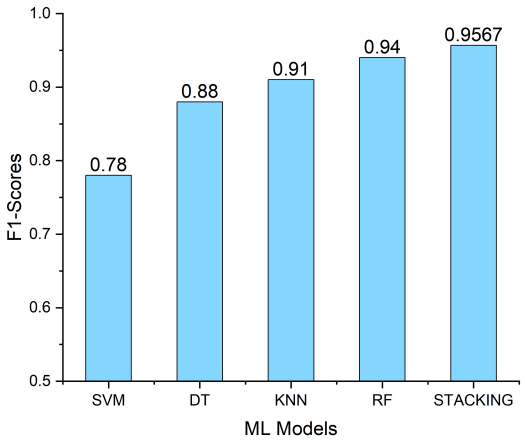


Fig. 2. Metrics of ADASYN sampling.

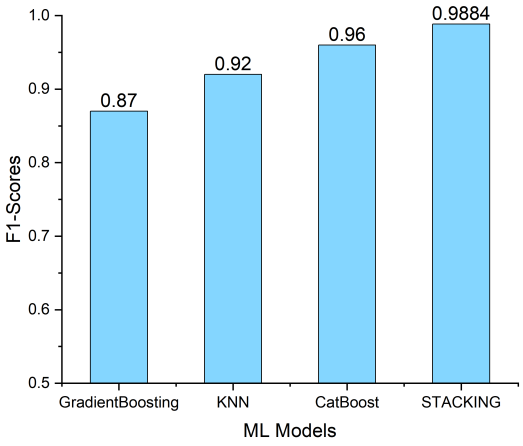


Fig. 3. Metrics of random oversampling.

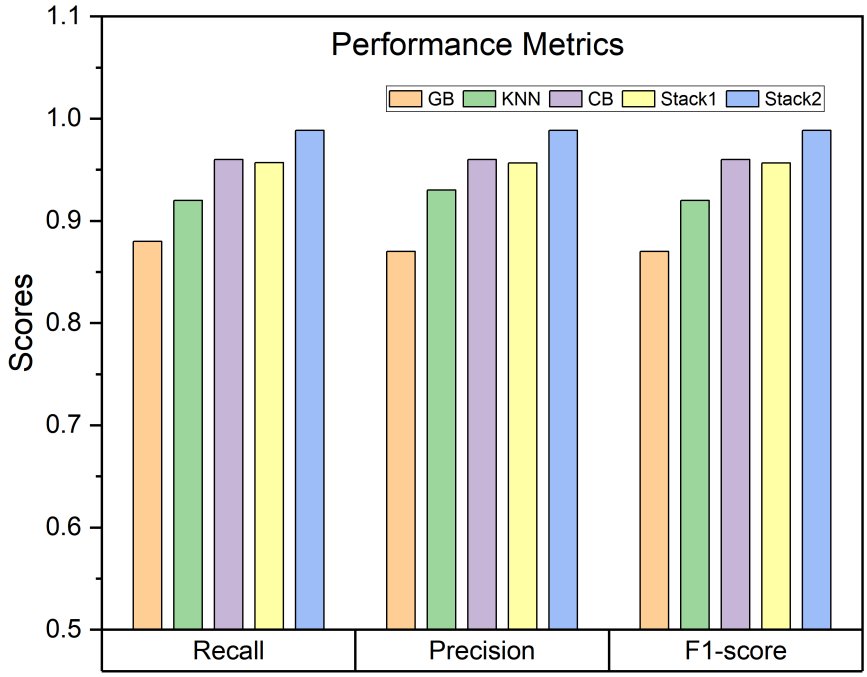


Fig. 4. Comparison of the results obtained by each stack.

Developing decision support systems is complex and aims to maximize performance measures like precision and recall to reduce false positives and negatives. While high and balanced values are ideal, clinical criteria, which can vary between hospitals, should determine which measure to prioritize. Therefore, hospital systems should be flexible, allowing clinicians to adapt protocols and prioritize specific performance measures, such as cesarean, vaginal, and assisted vaginal delivery decisions. The overall performance metrics of both stacks are represented in Table II with numerical values and in Fig. 4 through bar plots.

This study's findings exceed previous works aimed at predicting delivery methods and assessing cesarean risks, considering antepartum and intrapartum factors. In [3], a cohort of 25,038 patients with single pregnancies from the Service of

TABLE II. PERFORMANCE METRICS FOR DIFFERENT MODELS

Model	Recall	Precision	F1-Score
GradientBoosting	0.88	0.87	0.87
KNN	0.92	0.93	0.92
CatBoost	0.96	0.96	0.96
STACK1	0.957	0.9567	0.9567
STACK2	0.9885	0.9886	0.9884

Obstetrics and Gynaecology of the University Clinical Hospital evaluated the feasibility of using algorithms, namely, SVM, RF, and MultiLayer Perceptron, to predict the delivery method among cesarean, eutocic and assisted vaginal deliveries. The algorithms displayed an accuracy of 87%-90%. The study by Sultan et al. [17] and by Hasan et al. [26] on a sample containing 692 cesarean and 5465 non-cesarean samples collected from 4 hospitals in Spain worked on cesarean prediction. The SVC, XGB, and RF ensemble has achieved an F1-score of 96%. Concerning [27], a cohort of 13527 was prospectively assessed to predict cesarean deliveries. 32 classifiers were assessed, and the Quadratic discriminant analysis achieved an accuracy and F1-score of 97.9%.

The proposed KGC ensemble model consisting of KNN, Gradient Boosting, and CatBoost achieved a high F1 score of 98.84%, demonstrating its strong predictive capability. Gradient Boosting and CatBoost enhance the model's ability to handle complex, nonlinear relationships, while KNN adds the advantage of capturing local data patterns. The blend strikes a balance between bias and variance, in turn improving the model's generalization. However, the ensemble has challenges, including increased computational complexity and longer training times. Moreover, the decision-making process becomes less interpretable due to the complexity of Gradient Boosting and CatBoost, making it harder to understand how the model arrives at its predictions than KNN.

VI. CONCLUSION

An alarming increase in cesarean section rates surpassing WHO recommended levels poses significant medical, financial, and organizational challenges. Typically in low-risk pregnancies and non-elective cesarean sections, the possibility of vaginal or c-section deliveries relies on medical conditions and initiatives taken. Machine learning can act as a driving force and aid obstetricians in predicting the best feasible delivery mode based on the medical conditions, ensuring the mother's and newborn's safety. Machine learning can assist obstetricians working night shifts and in varying localities by providing real-time, data-driven insights and recommendations on the optimal delivery mode, accounting for factors such as limited staff availability, resource constraints, and patient demographics.

This research has proposed an optimized KGC classifier combining the algorithms of GB, KNN, and CatBoost along with a random oversampling balanced dataset to increase the predictive capability of the delivery mode. The performance of the proposed classifier has been tested against different performance metrics such as F-measure, recall and precision. Our proposed KGC classifier has achieved an F1-score of 98.84%, outperforming stack1 composed of KNN, DT, SVM, and RF.

In addition,KGC algorithm provides better results compared to previous studies. Hence we can conclude an optimized and random oversampled-balanced KGC classifier can reliably predict the C-section and vaginal classes.While our research has provided valuable insights,it is important to understand the dataset was limited to 6157 records. To further validate the model's robustness, future work is aimed to target the larger datasets with more intrapartum details. Incorporating such demographic variables in future studies could enhance the predictive power and applicability of the model. Further , our classifier can be applied to skin cancer, Parkinson's, gestational diabetes and socio-demographic data. In the future, we would like to work with deep learning algorithms for better predictive capabilities to implement a possible computer decision support system model to be built for the benefit of obstetricians and pregnant women.

ACKNOWLEDGMENT

The authors would like to thank ChatGPT for assisting in the proof reading.

REFERENCES

[1] Z. S. Lassi, T. Mansoor, R. A. Salam, J. K. Das, and Z. A. Bhutta, "Essential pre-pregnancy and pregnancy interventions for improved maternal, newborn and child health," *Reproductive health*, vol. 11, pp. 1–19, 2014.

[2] World Health Organization, "Medical eligibility criteria for contraceptive use." https://www.who.int/publications/i/item/WHO-RHR-15.02, 2015.

[3] A. D. R. Fernández, D. R. Fernández, and M. T. P. Sánchez, "Prediction of the mode of delivery using artificial intelligence algorithms," *Computer Methods and Programs in Biomedicine*, vol. 219, p. 106740, 2022.

[4] M. N. Islam, T. Mahmud, N. I. Khan, S. N. Mustafina, and A. N. Islam, "Exploring machine learning algorithms to find the best features for predicting modes of childbirth," *IEEE Access*, vol. 9, pp. 1680–1692, 2020.

[5] F. Parenting, "Different childbirth methods you must know." https://parenting.firstcry.com/articles/different-childbirth-methods-you-must-know/, 2020.

[6] L. Seshadri and G. Arjun, *Essentials of Obstetrics*. Wolters Kluwer India Pvt Ltd, 2015.

[7] M. S. Islam, M. A. Awal, J. N. Laboni, F. T. Pinki, S. Karmokar, K. M. Mumenin, S. Al-Ahmadi, M. A. Rahman, M. S. Hossain, and S. Mirjalili, "Hgsorf: Henry gas solubility optimization-based random forest for c-section prediction and xai-based cause analysis," *Computers in Biology and Medicine*, vol. 147, p. 105671, 2022.

[8] I. Khan and B. K. Khare, "Exploring the potential of machine learning in gynecological care: a review," *Archives of Gynecology and Obstetrics*, pp. 1–19, 2024.

[9] D. Lestari, F. I. Maulana, S. F. Persada, and P. D. P. Adi, "Machine learning for perinatal complication prediction: A systematic review," in *International Conference on Information, Communication and Computing Technology*, pp. 789–803, Springer, 2023.

[10] S. A. Tiruneh, T. T. T. Vu, D. L. Rolnik, H. J. Teede, and J. Enticott, "Machine learning algorithms versus classical regression models in pre-eclampsia prediction: A systematic review," *Current Hypertension Reports*, pp. 1–15, 2024.

[11] M. N. Islam, S. N. Mustafina, T. Mahmud, and N. I. Khan, "Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research agenda," *BMC pregnancy and childbirth*, vol. 22, no. 1, p. 348, 2022.

[12] K. Kolasa, B. Admassu, M. Holownia-Voloskova, K. J. Kedzior, J.-E. Poirrier, and S. Perni, "Systematic reviews of machine learning in healthcare: a literature review," *Expert Review of Pharmacoeconomics & Outcomes Research*, vol. 24, no. 1, pp. 63–115, 2024.

[13] J. Mas-Cabo, G. Prats-Boluda, J. Garcia-Casado, J. Alberola-Rubio, R. Monfort-Ortiz, C. Martinez-Saez, A. Perales, and Y. Ye-Lin, "Electrohysterogram for ann-based prediction of imminent labor in women with threatened preterm labor undergoing tocolytic therapy," *Sensors*, vol. 20, no. 9, p. 2681, 2020.

[14] C. Benalcazar-Parra, Y. Ye-Lin, J. Garcia-Casado, R. Monfort-Ortiz, J. Alberola-Rubio, A. Perales, and G. Prats-Boluda, "Prediction of labor induction success from the uterine electrohysterogram," *Journal of Sensors*, vol. 2019, no. 1, p. 6916251, 2019.

[15] E. Rohith, V. Sowmya, and K. Soman, "Convolutional neural networks for placenta cell classification," in *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, vol. 1, pp. 1273–1277, IEEE, 2019.

[16] D. S. Maylawati, M. A. Ramdhani, W. B. Zulfikar, I. Taufik, and W. Darmalaksana, "Expert system for predicting the early pregnancy with disorders using artificial neural network," in *2017 5th International Conference on Cyber and IT Service Management (CITSM)*, pp. 1–6, IEEE, 2017.

[17] N. Sultan, M. Hasan, M. F. Wahid, H. Saha, and A. Habib, "Cesarean section classification using machine learning with feature selection, data balancing, and explainability," *IEEE Access*, vol. 11, pp. 84487–84499, 2023.

[18] X. Bai, Z. Zhou, Z. Zheng, Y. Li, K. Liu, Y. Zheng, H. Yang, H. Zhu, S. Chen, and H. Pan, "Development and evaluation of machine learning models for predicting large-for-gestational-age newborns in women exposed to radiation prior to pregnancy," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, p. 174, 2024.

[19] Q.-Y. Yu, Y. Lin, Y.-R. Zhou, X.-J. Yang, and J. Hemelaar, "Predicting risk of preterm birth in singleton pregnancies using machine learning algorithms," *Frontiers in big Data*, vol. 7, p. 1291196, 2024.

[20] Y. Zhang, S. Lu, Y. Wu, W. Hu, and Z. Yuan, "The prediction of preterm birth using time-series technology-based machine learning: retrospective cohort study," *JMIR Medical Informatics*, vol. 10, no. 6, p. e33835, 2022.

[21] G. Obaido, O. Achilonu, B. Ogbuokiri, C. S. Amadi, L. Habeebullahi, T. Ohalloran, C. W. Chukwu, E. Mienye, M. Aliyu, O. Fasawe, *et al.*, "An improved framework for detecting thyroid disease using filter-based feature selection and stacking ensemble," *IEEE Access*, 2024.

[22] P. Sindhu, G. S. Peruri, and M. Yalavarthi, "An empirically based object-oriented testing using machine learning," *EAI Endorsed Transactions on Internet of Things*, vol. 10, 2024.

[23] N. P, I. S. U, B. V, and H. R, "Preterm birth prediction using machine learning," in *2024 International Conference on Inventive Computation Technologies (ICICT)*, pp. 1–6, 2024.

[24] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, 2020.

[25] C. Campillo-Artero, M. Serra-Burriel, and A. Calvo-Pérez, "Predictive modeling of emergency cesarean delivery," *PloS one*, vol. 13, no. 1, p. e0191248, 2018.

[26] M. Hasan, M. J. Zobair, S. Akter, M. Ashef, N. Akter, and N. B. Sadia, "Ensemble based machine learning model for early detection of mother's delivery mode," in *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1–6, 2023.

[27] M. Kowsher, N. J. Prottasha, A. Tahabilder, H. Kaiser, M. Abdur-Rakib, and M. S. Alam, "Predicting the appropriate mode of childbirth using machine learning algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, 2021.

MH-LViT: Multi-path Hybrid Lightweight ViT Models with Enhancement Training

Yating Li[1], Wenwu He[2], Shuli Xing[3], Hengliang Zhu[4]
School of Computer Science and Mathematics, Fujian University of Technology, Fuzhou 350118, China[1,2,3,4]
Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fuzhou 350118, China[2,3,4]

***Abstract*—Vision Transformers (ViTs) have become increasingly popular in various vision tasks. However, it also becomes challenging to adapt them to applications where computation resources are very limited. To this end, we propose a novel multi-path hybrid architecture and develop a series of lightweight ViT (MH-LViT) models to balance well performance and complexity. Specifically, a triple-path architecture is exploited to facilitate feature representation learning that divides and shuffles image features in channels following a feature scale balancing strategy. In the first path ViTs are utilized to extract global features while in the second path CNNs are introduced to focus more on local features extraction. The third path completes the representation learning with a residual connection. Based on the developed lightweight models, a novel knowledge distillation framework IntPNKD (Normalized Knowledge Distillation with Intermediate Layer Prediction Alignment) is proposed to enhance their representation ability, and in the meanwhile, an additional Mixup regularization term is introduced to further improve their generalization ability. Experimental results on benchmark datasets show that, with the multi-path architecture, the developed lightweight models perform well by utilizing existing CNN and ViT components, and with the proposed model enhancement training methods, the resultant models outperform notably their competitors. For example, on dataset miniImageNet, our MH-LViT_M3 improves the top-1 accuracy by 4.43% and runs 4x faster on GPU, compared with EdgeViT-S; on dataset CIFA10, our MH-LViT_M1 improves the top-1 accuracy by 1.24% and the enhanced version MH-LViT_M1* by 2.28%, compared to the recent model EfficientViT_M1.**

***Keywords*—Multi-path hybrid; lightweight ViT; normalized knowledge distillation; Mixup regularization**

I. Introduction

In recent years, Vision transformers (ViTs) [1] have received increasing attention in many visual tasks, achieving remarkable results in tasks such as image classification [1], [2], [3], target detection [4], [5], [6], [7] and semantic segmentation [8], [9], [10], [11], [12], [13]. However, the computational overhead of the self-attention mechanism makes ViTs less efficient than Convolutional Neural Networks (CNNs) [14] on memory and computationally constrained devices. The huge model size and computational cost make it challenging to adapt them to real-time applications. Therefore, researchers tend to build lightweight and efficient ViT models.

Some approaches aim to build lightweight versions of ViT models by reducing the number of feature channels or self-attentive heads. Nevertheless, such approaches usually lead to significant performance degradation. The author in [15] successfully improves the performance of existing tiny ViTs by introducing a plugin that groups and shuffles feature channels. Alternatively, some researchers have attempted to improve model performance through specific designs, such as combining computationally expensive self-attention with efficient convolutional operations to create hybrid efficient ViTs [16], [17], [18], [19], [20]. Among them, MobileViT [16] combines the image-specific inductive bias of CNNs and the global information processing capability of ViTs to encode both local and global information efficiently. EdgeViT [19] combines the attention mechanism and CNNs through the implementation of local-global-local (LGL) blocks. Nevertheless, most of these methods serially stack self-attention and convolutional layers, and the extraction of global features often compromises previously extracted local features, failing to take full advantage of global and local features. To overcome this problem, several studies have begun to explore the application of parallel structures in feature extraction. The parallel structure allows the self-attention and convolutional layers to work simultaneously and extracts global and local features independently, and then fuses them in some way. This structure can maintain the integrity of local features while combining global features to achieve a more comprehensive feature representation. For example, TransXNet [20] efficiently extracts and fuses global and local features by combining self-attention and convolution in parallel to achieve excellent performance.

Therefore, this paper adopts a parallel structure approach to extract both global and local features, fully exploiting the diversity of features to enhance the model performance. Unlike previous parallel structures, we have added an additional residual branch to enhance the learning ability of feature representation, without explicitly increasing parameters or computational overhead. Specifically, we exploit a feature scale balancing strategy to divide input features into three parts. The first one is fed into the Transformer branch to extract global features, the second one is fed into the CNN branch to extract local features, and the last one is directly used to form the Residual branch. Subsequently, the feature fusion module is employed to shuffle extracted features and balance their chances to be processed in different branches. Based on this, we conclude that the tiny model has strong features representation ability. To further release their representation potential, we propose a new knowledge distillation framework, IntPNKD (Normalized Knowledge Distillation with Intermediate Layer Prediction Alignment), to effectively improve the inference performance without additional inference cost. In addition to the standard knowledge distillation procedure of NKD (Normalized Knowledge Distillation) [21], IntPNKD aligns the predictions made respectively on intermediate feature maps of the teacher and the student. In the meanwhile, an image mixing regularizer

is introduced to further enhance the generalization ability of the resultant model. The main contributions of this paper are summarized as follows:

- We propose a multi-path hybrid architecture to design lightweight ViT model (MH-LViT), where a triple-branch architecture (transformer, CNN, and residual) is employed to extract efficiently local and global features, and the features extracted by multiple branches are further shuffled and re-assigned with a feature fusion module.
- We propose a new knowledge distillation framework IntPNKD to improve the representation ability of MH-LViT, which is further enhanced by introducing an image mixing regularization term. The achieved performance improvement costs nothing in inference.
- We develop a series MH-LViT models with various sizes that are tested on multiple benchmark datasets. The experimental results show that our models balance well the efficiency and the accuracy.

II. Related Work

A. Efficient Vision Transformers

In order to reduce the number of parameters and the computational overhead, a series of works on efficient vision transformers have been proposed, which cover a wide range of interesting ideas such as lightweight module design, model compression, token compression, hybrid model design, and so on. Window-based Self-Attention ViT proposed in Swin [4] reduces the computation cost of each transformer block by dividing the feature map into windows and restricting the attention operation within a local window. In this way, the length of input sequence fed into Transformer is reduced, leading to improved efficiency. Wang et al. [22] proposed a spatial-reduction attention (SRA) to reduce the computation cost of self-attention, by downsampling the spatial resolution dimensions of the input Query and Key branches via a lightweight depthwise convolution. Token merging [23] achieves parameter compression by merging tokens with large semantic similarity. CVT [24] and LeViT [17] insert some convolutional layers into Transformer layers, to downsample feature maps and perform local information fusion, where the Transformer layers are used to capture the global information from deep features. Liu et al. [25] proposed a cascade group attention module that uses only three or four attention heads, to increase the diversity of features while reducing computational redundancy in multi-head attention.

B. Efficient Convolutional Neural Networks

Actually, before the arrival of ViT, models based on Convolutional Neural Networks (CNNs) faced the same challenge to deploy them to devices with limited computational resources. To fix this, a series of elegant work have been proposed. MobileNets [26], [27] leverage depthwise separable convolutions to reduce the computation complexity. ShuffleNet [28] utilizes pointwise group convolution and performs channel shuffle operations to reduce the computation cost. IDConv[20] reduces the computation overhead by utilizing dynamic deep convolutions and adaptive average pooling. Overall, efficient CNNs work well in a variety of scenarios which inspires us to develop a novel model that combines efficient CNNs and ViT in some way, to extract effectively both global and local features.

C. Knowledge Distillation

Knowledge distillation (KD) has attracted wide attention in the field of model compression, which typically transfers knowledge from a teacher (model) to a student (model) to improve the latter's performance. This framework, originally proposed by Hinton et al. [29], utilizes both the hard labels of ground truth and the soft ones provided by the teacher to guide the learning process of the student. Recently, there have emerged studies [30], [21] focused on KD in ViT. For instance, DeiT [30] introduces a novel distillation procedure which relies on a distillation token ensuring that the student learns from the teacher through attention, to achieve efficient transformers with competitive top-1 accuracy on ImageNet. MixSKD [31] is a self-knowledge distillation framework that enables the network to learn cross-image knowledge by modeling supervisory signals from mixup images. The most relevant work on knowledge distillation to ours is NKD [21], which uses cross-entropy to align respectively the probability distributions of target class and non-target classes. We build our KD framework on NKD and introduce Middle Feature Prediction Alignment into it for better knowledge transfer.

III. Preliminaries

A. EfficientViT

Unlike the classic ViT [1], EfficientViT [25] adopts a sandwich layout and a cascade group attention. Specifically, the cascade group self-attention layer Φ_i^A is sandwiched between two FFN (Feed Forward Network) layers each of them is denoted with Φ_i^F. The operation of the i-th block can be formulated as:

$$X_{i+1} = \Phi_i^{\mathrm{F}}(\Phi_i^{\mathrm{A}}(\Phi_i^{\mathrm{F}}(X_i))), \tag{1}$$

where X_i is the input feature of the i-th block. In the the cascade group self-attention layer Φ_i^A, the input feature X_i is divided into J parts which are fed into J attention heads, where the output of each head is added to the subsequent heads to enrich the feature information. Formally, the operation of the j-th attention head can be expressed as:

$$\mathring{X}_i^j = \begin{cases} \mathrm{Attn}(X_i^j), & if\ j = 0, \\ \mathrm{Attn}(X_i^j + \mathrm{Attn}(\mathring{X}_i^{j-1})), & if\ j \geq 1, \end{cases} \tag{2}$$

where, X_i^j denotes the j-th split of input X_i and $\mathring{X}_i^j$ is its corresponding attention output. Then J attention outputs are concatenated:

$$\hat{X}_i = \Phi_i^A(X_i) = \mathrm{concat}[\{\tilde{X}_i^j\}_{j=1}^J], \tag{3}$$

where, $\hat{X}_i$ is the output of the i-th cascade group attention layer.

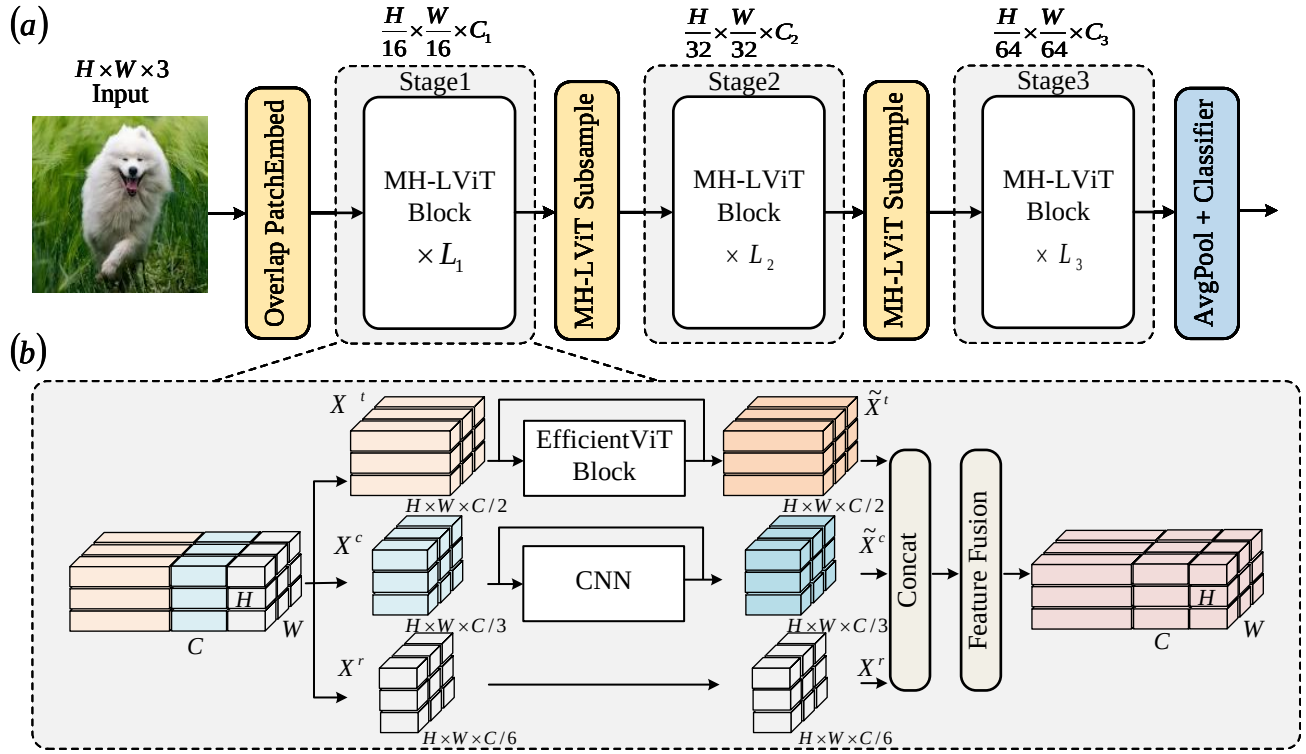


Fig. 1. Architecture of MH-LViT: (a) Overall architecture of MH-LViT; (b) Multi-Path branch block.

B. Normalized Knowledge Distillation

Normalized Knowledge Distillation (NKD) [21] is an improvement version of the decoupled knowledge distillation (DKD) [32] that normalizes the non-target logits and utilizes Cross-Entropy (CE) instead of Kullback-Leibler (KL) divergence to align the target and non-target class distributions respectively. In particular, the loss of NKD can be formulated as follows:

$$L_{NKD} = -T_t log(S_t) - \gamma \cdot \lambda^2 \cdot \sum_{k \neq t} \mathcal{N}(T_k^{\lambda}) log(\mathcal{N}(S_k^{\lambda})), \quad (4)$$

where, the index t denotes the target class, T_k (S_k) denotes the output probabilities of teacher (student) model corresponding to the k-th class, λ is the KD temperature [29], γ is the hyper-parameter to balance the two items in the loss, and $\mathcal{N}(*)$ denote the normalization operation.

C. Mixup

Mixup [33] generates a mixed image (x_{ab}, y_{ab}) by linearly combining a pair of original images $\{x_a, y_a; x_b, y_b\} \in D$, where D denote a data set, x_a (x_b) denotes the image a (b) and y_a (y_b) is its corresponding label. The mixed image and its corresponding label are formulated as follows:

$$\begin{aligned} x_{ab} &= \lambda x_a + (1-\lambda) x_b, \\ y_{ab} &= \lambda y_a + (1-\lambda) y_b, \end{aligned} \quad (5)$$

where, the combination is controlled by the mixing factor λ sampled from the beta distribution.

IV. Methods

A. Overview

We propose a new multi-path hybrid architecture to design lightweight ViT model MH-LViT. The overall architecture of MH-ViT is shown in Fig. 1(a). The model adopts a hierarchical architecture, that effectively reduces the resolution of feature maps during the forward propagation process, while gradually increases the number of channels for feature map. Specifically, MH-LViT contains three stages, each of which consists of L MH-LViT blocks. In order to reduce the amount of parameters and computation overhead, we divide the input features by channels in the ratio of $3:2:1$ via the feature scale balancing strategy, and then learn images' representation in parallel through three paths, i.e. Transformer, CNN and Residual. By introducing the multi-hybrid structure, the model can capture global information and local details at the same time, to understand well the images for latter computer vision (CV) tasks. In addition, a new type of knowledge distillation and a mixup regularization are exploited to further improve the inference performance without any additional inference cost.

B. Lightweight Model Design

1) Multi-Path branch: As shown in Fig. 1(b), the MH-LViT Block is mainly composed of a three-branch architecture

and a feature fusion module. In particular, the three branches includes an efficient Transformer (EfficientViT Block), a lightweight CNN (IDConv) and a residual connection. Formally, the operations in MH-LViT Block can be written as follows:

$$X_i^t, X_i^c, X_i^r = \text{Split}(X_i),\ X_i^t : X_i^c : X_i^r = 3:2:1; \quad (6)$$

$$\tilde{X}_i^t = \text{EfficientViTBlock}(X_i^t); \quad (7)$$

$$\tilde{X}_i^c = \text{IDConv}(X_i^c); \quad (8)$$

$$\tilde{X}_i = \text{Concat}(\tilde{X}_i^t, \tilde{X}_i^c, X_i^r); \quad (9)$$

$$X_{i+1} = \text{FF}(\tilde{X}_i). \quad (10)$$

Here, with a little notation abuse, X_i denotes the input feature of the i-th MH-LViT block and X_{i+1} the corresponding output. $Split(\cdot)$ denotes the operation to divide the input feature into three splits with the specified ratio and : denotes the ratio of the number of channels.

2) Transformer branch: It is particularly important to capture effective global information of images for CV tasks. To this end, we introduce the cutting-edge EfficientViT [25] to our model to extract global features. EfficientViT not only inherits powerful capabilities of the vanilla ViT model, but also significantly improves the computational efficiency via a sandwich layout and a cascade group attention. In addition, the depthwise convolution (DWConv) is used for information fusion before the final FFN layer of the block. The part of input features X_i^t is fed into the EfficientViT branch, to learn the representations of images base on the global information. The processing of X_i^t in an EfficientViT Block can be expressed as follows:

$$\tilde{X}_i^t = \text{FFN}(\text{DW}((\Phi_i^A(\text{FFN}(\text{DW}(X_i^t)))))), \quad (11)$$

where, $DW(\cdot)$ denotes the depthwise convolution and other notations denote the same operation as indicated before.

3) CNN branch: In order to inject inductive bias for local feature extraction, this paper introduces IDConv (Input-dependent Depthwise Convolution) [20] to extract image local information. IDConv can dynamically generate convolution kernels, which enhances the adaptability and characterization of CNN for different data features. Firstly, the spatial dimension of input features $X_i^c \in C/3{\times}H{\times}W$ is compressed to K^2 by using adaptive pooling, to aggregate the spatial context information. Subsequently, two consecutive $1{\times}1$ convolutional layers are utilized to generate the attention map $A' \in (G{\times}C/3){\times}K^2$, where G represents the number of attention groups. Then, A' is reshaped to $G{\times}C/3{\times}K^2$, where a softmax operation is applied in the G-dimension to generate the attention weights $A \in G{\times}C/3{\times}K^2$. Finally, the attention weight A is element-wise multiplied with a set of learnable parameters $P \in G{\times}C/3{\times}K^2$ and summed along the G-dimension to obtain the input-dependent deep convolution kernel $W \in C/3{\times}K^2$. This whole process can be expressed as follows:

$$A_i' = \text{Conv}(\text{Conv}(\text{AdaptivePool}(X_i^c))); \quad (12)$$

$$A_i = \text{Softmax}(\text{Reshape}(A_i')); \quad (13)$$

$$W_i = \sum_{g=1}^{G} (P_i)_g (A_i)_g; \quad (14)$$

$$\tilde{X}_i^c = W_i X_i^c. \quad (15)$$

4) Residual branch: As we know, in CV models, the information of original features may be weakened or lost when a series of transformation operations have been performed on them. To further enrich the feature representation and to ensure that the model can make full use of the original information, we preserve part of original features, i.e. X_i^r, and do not perform any transforming or processing on these features. In this way, we include both processed and unprocessed raw features to enhance the learning ability of feature representation, without explicit increasing in parameters or computation overhead.

5) Fusion module: For the features extracted through multiple paths, we introduce a feature fusion module to shuffle them, to let features in each branch get the chance to be fed into other branches. The feature fusion module mainly implements channel shuffle, aiming to shuffle the limited interaction between branches by shuffling features from different branches. This mixes the features from different branches, thereby achieving information fusion. Specifically, the global features extracted through Transformer Branch, the local features through CNN Branch, and the features for residual connection are first concatenated and then input into the fusion module, where a channel shuffle operation is utilized to shuffle inputs. In particular, the input feature tensor $\tilde{X}_i$ is grouped into a certain number of groups by the channel-dimension, and then the channels within each group are rearranged to obtain the shuffled tensor $X_i^{shuffle}$. Finally, $X_i^{shuffle}$ is restored to the original shape to obtain the fused feature tensor X_{i+1}. This module effectively mixes the features extracted from different branches, to avoid features in one branch are locked within this branch. Formally, the operations in FF module can be written as follows:

$$\tilde{X}_i^{group} = \text{Split}_{\text{group}}(\tilde{X}_i); \quad (16)$$

$$\tilde{X}_i^{shuffle} = \text{Shuffle}(\tilde{X}_i^{group}); \quad (17)$$

$$X_{i+1} = \text{Reshape}(\tilde{X}_i^{shuffle}). \quad (18)$$

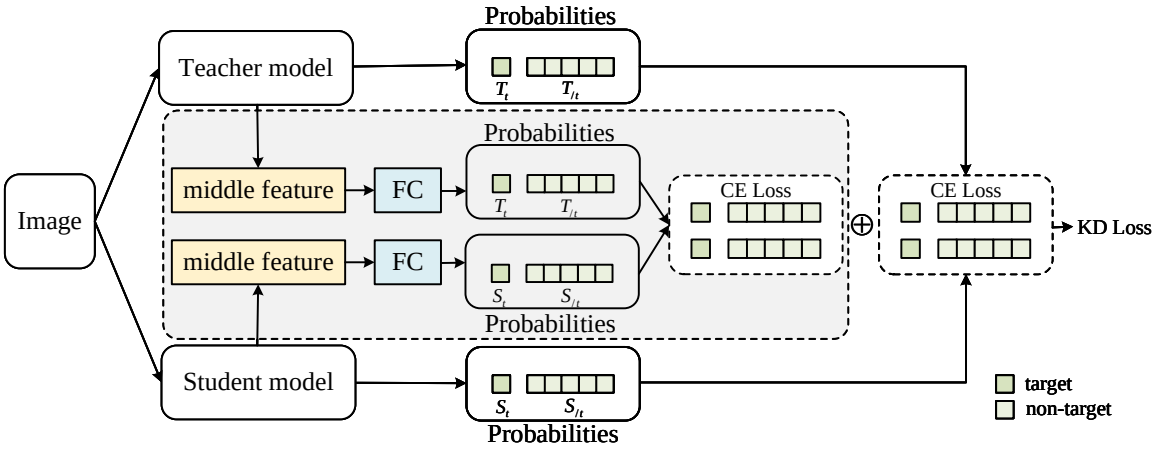


Fig. 2. Illustration of IntPNKD (Normalized knowledge distillation with intermediate layer prediction alignment).

C. Lightweight Model Enhancements

1) IntPNKD: Limited to its lightweight, the developed tiny model inevitably sacrifice its performance to some degree. To this end, we further propose a new distillation framework IntPNKD to improve its representation ability. Specifically, as shown in Fig. 2, IntPNKD transfers knowledge from the teacher (standard model) to the student (lightweight model), by aligning not only the output logits of two models (as vanilla NKD does) but also the predictions made on intermediate layers' feature maps. While the traditional knowledge distillation (NKD) focuses on aligning the final output logits, it may overlook important feature information encoded in the intermediate layers, which plays a critical role in the overall learning process. With the additional alignment, IntPNKD improves the effectiveness of knowledge distillation from teacher to student. Formally, let X^T_{mid} (X^S_{mid}) denote the features of the intermediate layer of teacher (student) model, $\text{FC}(\cdot)$ denote the fully connected layer and $\text{P}(\cdot)$ the softmax operation. Then the total distillation loss of IntPNKD can be written as:

$$L_{IntPNKD} = L_{NKD} + \text{KL}(\text{P}(\text{FC}(X^T_{mid})), \text{P}(\text{FC}(X^S_{mid}))). \tag{19}$$

2) Mixup regularization: Following the idea of image mixing in MixSKD [31], we introduce a regularizer L_{MR} to our case (the tiny student model), to enhance the generalization ability of the developed model. As shown in Fig. 3, we mix randomly two original images, e.g. x_a and x_b, to obtain a mixup one x_{ab}, and expect that the probability distribution output by the model on x_{ab} and the one given by mixing the logits of x_a and x_b are not too far from each other. In particular, the KL divergence is utilized to form the regularization term, to guide the model make relatively stable predictions on mixup image and original ones, leading to improved inference performance. The regularization term can be formulated as:

$$L_{MR} = \text{KL}(\text{P}^{\text{s}}(X_a, X_b), \text{P}^{\text{s}}(X_{ab})), \tag{20}$$

$$\text{P}^{\text{S}}(X_a, X_b) = \text{Softmax}(\lambda \text{S}(X_a) + (1-\lambda)\text{S}(X_b)), \tag{21}$$

$$\text{P}^{\text{S}}(X_{ab}) = \text{Softmax}(\text{S}(X_{ab})), \tag{22}$$

where, $\text{P}(\cdot)$ denotes the softmax operation, $S(\cdot)$ the logit (before softmax operation) output by the tiny student model, and λ is the mixing factor as in Eq. (5).

Combining everything together, the total loss function L used to train the lightweight model can be written as:

$$L = L_{CE} + L_{IntPNKD} + L_{MR}, \tag{23}$$

where, L_{CE} is the cross-entropy loss guided by the ground truth.

V. Experiments

In this section, we perform experiments on several benchmark datasets, to validate the effectiveness of proposed methods. We first elaborate on the implementation details of the experiments and then present the main experimental results of the proposed models and the relevant baseline models, which are discussed in depth. To further dissect the performance of proposed model, we also conduct a series of ablation studies to evaluate the practical effects of its key components.

A. Implementation Details

We conduct image classification experiments on three benchmark datasets, i.e. CIFAR10 [34], CIFAR100 [34] and miniImageNet [35]. In building the models, we used two tool libraries, i.e. PyTorch 1.11.0 [36] and Timm 0.5.4 [37]. The AdamW [38] optimizer and the cosine learning rate scheduler are used to train related models, each of which is trained 300 epochs from scratch on an Nvidia A100 GPU. For the input images, we resize and randomly crop them to 224 × 224 pixels.

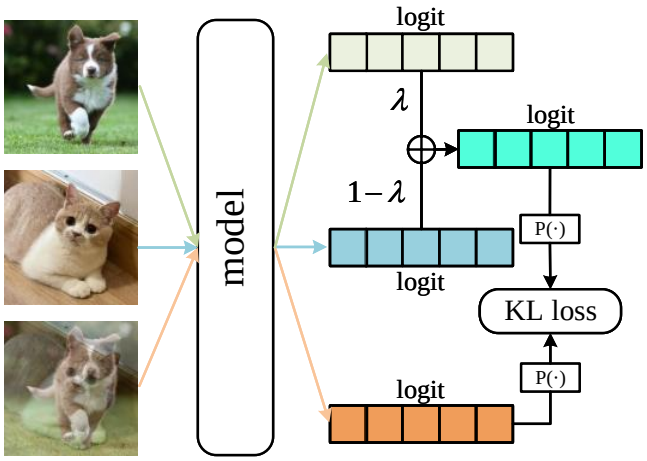


Fig. 3. Illustration of mixup regularization.

The batch size is fixed to 64, the initial learning rate is set to 1×10^{-3}, and the weight decay is 0.01. In addition, we use the same data augmentation strategies as in[30], including auto-augmentation [39] and random erasing [40]. We evaluate the model with the size, the top-1 accuracy, the throughput and Flops to get a full picture of its performance. As for KD, we utilize recently proposed CARTE-B [41] as the teacher model, which performs well with a medium model size by pre-training on ImageNet.

B. Main Results

We compare the developed MH-LViT with popular efficient models based on CNN or ViT and report the results. The results show that MH-LViT models with different sizes achieve the best accuracy and speed tradeoff across benchmark datasets in most cases.

1) Results on miniImageNet: Table I summarizes the results on the dataset miniImageNet achieved by the proposed lightweight models and multiple SOTA competitors. We first compare MH-LViT with tiny models whose size are close to it. As can be seen from the table, MH-LViT_M1 outperforms EfficientViT_M1 and EdgeViT-XXS by up to 5.57% and 6.07% respectively, with a comparable number of parameters. Compared with ShuffleNetV2 2.0x, MH-LViT_M2 improves the top-1 accuracy by 1.46% with fewer parameters and faster inference speed. Compared with EdgeViT-S, MH-LViT_M3 improves the top-1 accuracy by 1.74% and runs 4x faster on GPU. More interestingly, the enhanced version MH-LViT* outperforms all competitors in terms of accuracy. We can observe from Table I that, MH-LViT_M3* achieves 81.50% top-1accuracy on this dataset, which outperforms EdgeViT-S by up to 7.41% with a comparable number of parameters. Compared with EdgeViT-S, MH-LViT_M2* improves the top-1 accuracy by 4.19% with lower parameters and Flops. MH-ViT with various model sizes perform well on dataset miniImageNet and the results show that the multiple-path design is effective and model enhancement training strategy even goes an extra mile.

2) Results on CIFAR100: Table II shows in detail the performance of proposed models and its competitors on dataset CIFAR-100. For example, MH-LViT_M1 improves the top-1 accuracy by 2.93% compared to EfficientViT_M1. MH-LViT_M2 improves the top-1 accuracy by 2.32% and runs 3x faster on the GPU, compared to EdgeViT-XS that has a comparable number of parameters. Compared to ShuffleNetV2 2.0x that achieves the best top-1 accuracy on this dataset, MH-LViT_M2 provides a competitive top-1 accuracy but has a 1.8x throughput. MH-LViT_M2* even goes further and can beat ShuffleNetV2 2.0x with 4.01% improvement in top-1 accuracy. Compared to models with higher throughput such as EfficientViT_M3, MH-LViT_M2 improves its top-1 accuracy by 3.32% while maintaining a similar throughput. When model size goes large, MH-LViT_M3* outperforms EdgeViT-S in top-1 accuracy by 5.54% and runs 4.5x faster than it on GPU.

3) Results on CIFAR10: Table III summarizes experimental results of related models on dataset CIFA10. As what found in Tables I and II, from Table III one can find similar performance advantages of MH-LViT models over their competitors. For example, MH-LViT_M1 runs 5.2x faster than EdgeViT-XXS while improves the top-1 accuracy by 0.39%. The enhanced version MH-LViT_M1* achieves up to 96.28% top-1 accuracy on CIFA10 with 3.1M parameters. Compared with EdgeViT-XS model, MH-LViT_M2* runs 3.7x faster on the GPU with a comparable number of parameters and top-1 accuracy. MH-LViT_M3* achieves the highest top-1 accuracy of 96.75% with 12.5M parameters.

Overall, experimental results on bench mark datasets validate that the developed lightweight models with multiple-path design work well and can strike a good balance between the efficiency and the accuracy. The proposed model enhancement training strategy is effective and can provide further significant accuracy improvement without any additional inference cost.

C. Ablation Study

In this section, we validate the effectiveness of main components of MH-LViT/MH-LViT*, such as Multi-Path Branch, feature fusion module, IntPNKD, and mixup regularization, by

TABLE I. MH-LVIT IMAGE CLASSIFICATION PERFORMANCE ON MINIIMAGENET WITH COMPARISONS TO STATE-OF-THE-ART EFFICIENT CNN AND VIT

Model	Params(M)↓	ACC_Top1 (%)↑	Throughput(images/s)↑	Flops(M)↓	Input	Epoch
ShuffleNetV2 0.5x	1.3	61.58	34012	44	224	300
MobileViT-XXS	1.3	45.66	4456	273	224	300
ShuffleNetV2 1.0x	2.3	73.19	5454	152	224	300
MobileViT-XS	2.3	55.12	3344	744	224	300
MobileNetV3-Small	2.5	69.82	9031	65	224	300
EfficientViT_M1	3.0	69.96	20093	167	224	300
MH-LViT_M1	3.1	75.53	19126	130	224	300
MH-LViT_M1*	3.1	**76.44**	19126	130	224	300
EdgeViT-XXS	4.1	69.46	3638	546	224	300
MobileNetV3-Large	5.4	74.01	7920	271	224	300
MobileViT-S	5.6	73.59	1939	1464	224	300
EdgeViT-XS	6.7	73.20	3852	1123	224	300
MH-LViT_M2	6.7	76.76	14325	377	224	300
MH-LViT_M2*	6.7	**78.28**	14325	377	224	300
EfficientViT_M3	6.9	68.87	16644	263	224	300
ShuffleNetV2 2.0x	7.4	75.30	7540	596	224	300
LeViT-128	9.2	65.42	10905	371	224	300
LeViT-192	10.9	67.65	8837	605	224	300
EdgeViT-S	11.1	74.09	2274	1897	224	300
EfficientViT_M5	12.4	70.98	10621	522	224	300
MH-LViT_M3	12.5	78.52	10347	452	224	300
MH-LViT_M3*	12.5	**81.50**	10347	452	224	300
LeViT-256	18.9	68.88	6494	1059	224	300
LeViT-384	39.1	69.45	3883	2250	224	300

TABLE II. MH-LVIT IMAGE CLASSIFICATION PERFORMANCE ON CIFAR100 WITH COMPARISONS TO STATE-OF-THE-ART EFFICIENT CNN AND VIT

Model	Params(M)↓	ACC_Top1 (%)↑	Throughput(images/s)↑	Flops(M)↓	Input	Epoch
ShuffleNetV2 0.5x	1.3	72.16	34012	44	224	300
MobileViT-XXS	1.3	64.38	4456	273	224	300
ShuffleNetV2 1.0x	2.3	76.19	5454	152	224	300
MobileViT-XS	2.3	75.21	3344	744	224	300
MobileNetV3-Small	2.5	71.21	9031	65	224	300
EfficientViT_M1	3.0	73.46	20093	167	224	300
MH-LViT_M1	3.1	76.39	19126	130	224	300
MH-LViT_M1*	3.1	**80.19**	19126	130	224	300
EdgeViT-XXS	4.1	66.79	3638	546	224	300
MobileNetV3-Large	5.4	71.74	7920	271	224	300
MobileViT-S	5.6	75.92	1939	1464	224	300
EdgeViT-XS	6.7	75.45	3852	1123	224	300
MH-LViT_M2	6.7	77.80	14325	377	224	300
MH-LViT_M2*	6.7	**81.78**	14325	377	224	300
EfficientViT_M3	6.9	74.48	16644	263	224	300
ShuffleNetV2 2.0x	7.4	77.77	7540	596	224	300
LeViT-128	9.2	69.17	10905	371	224	300
LeViT-192	10.9	71.08	8837	605	224	300
EdgeViT-S	11.1	76.61	2274	1897	224	300
EfficientViT_M5	12.4	74.24	10621	522	224	300
MH-LViT_M3	12.5	78.35	10347	452	224	300
MH-LViT_M3*	12.5	**82.15**	10347	452	224	300
LeViT-256	18.9	71.14	6494	1059	224	300
LeViT-384	39.1	72.17	3883	2250	224	300

performing an ablation study on the dataset CIFAR100. The experimental results are summarized in Tables IV to VI.

1) Multi-Path branch: As shown in Table IV, we remove respectively the CNN Branch, the Transformer Branch, and the Residual Branch but keep the same model size as the complete one, to verify their effectiveness. As can be seen from the table, when the CNN Branch is removed, top-1 accuracy of MH-LViT_M1 decreases by 1.56%, which shows that CNNs in MH-ViT is helpful to enhance the leaning representation ability. When the Transformer Branch is removed, the top-1 accuracy decreases by 4.43%, which validates well that Transformer blocks are crucial for our model. Meanwhile, the Residual Branch also plays an auxiliary role in improving the proposed model, and when it is removed, the model accuracy decreases by 0.55%.

2) Feature fusion module: The ablation experimental result of Feature Fusion Module is also reported in Table IV, where we can observe a decrease of 0.9% in top-1 accuracy when FF module is removed from MH-LViT_M1. This result reveals the important auxiliary role of FF module in the developed model.

TABLE III. MH-LVITIMAGE CLASSIFICATION PERFORMANCE ON CIFAR10 WITH COMPARISONS TO STATE-OF-THE-ART EFFICIENT CNN AND VIT

Model	Params(M)↓	ACC_Top1 (%)↑	Throughput(images/s)↑	Flops(M)↓	Input	Epoch
ShuffleNetV2 0.5x	1.3	91.84	34012	44	224	300
MobileViT-XXS	1.3	94.57	4456	273	224	300
ShuffleNetV2 1.0x	2.3	94.50	5454	152	224	300
MobileViT-XS	2.3	88.58	3344	744	224	300
MobileNetV3-Small	2.5	94.66	9031	65	224	300
EfficientViT_M1	3.0	94.0	20093	167	224	300
MH-LViT_M1	3.1	95.24	19126	130	224	300
MH-LViT_M1*	3.1	**96.28**	19126	130	224	300
EdgeViT-XXS	4.1	94.85	3638	546	224	300
MobileNetV3-Large	5.4	95.56	7920	271	224	300
MobileViT-S	5.6	95.11	1939	1464	224	300
EdgeViT-XS	6.7	95.52	3852	1123	224	300
MH-LViT_M2	6.7	95.57	14325	377	224	300
MH-LViT_M2*	6.7	**96.67**	14325	377	224	300
EfficientViT_M3	6.9	94.51	16644	263	224	300
ShuffleNetV2 2.0x	7.4	95.15	7540	596	224	300
LeViT-128	9.2	94.0	10905	371	224	300
LeViT-192	10.9	94.33	8837	605	224	300
EdgeViT-S	11.1	95.58	2274	1897	224	300
EfficientViT_M5	12.4	94.61	10621	522	224	300
MH-LViT_M3	12.5	95.78	10347	452	224	300
MH-LViT_M3*	12.5	**96.75**	10347	452	224	300
LeViT-256	18.9	94.36	6494	1059	224	300
LeViT-384	39.1	94.59	3883	2250	224	300

Feature fusion can effectively integrate the features extracted from different branches to avoid features in one branch are locked within this branch and get a chance to be processed through other branches, enabling the model to understand input images well which in turn improves the classification accuracy. This indicates that feature fusion is an important factor in improving the performance of MH-LViT.

3) Feature scale balancing strategy: In the proposed models we divide the input features by channels in a ratio of $3:2:1$ via the feature scale balancing strategy, to learn image representation in parallel through multiple paths. In this section, we vary this ratio to investigate its effects on model performance. As shown in Table V, we split the features with several typical ratios, where the baseline model EfficientViT_M1 and a double-paths scenario with the ratio of $1:1$ are also included. From the table one can observe that, compare to the baseline, the model utilizing double-paths with ratio $1:1$ improves the top-1 accuracy significantly, while the model using the ratio of $1:1:1$ is 0.4% higher than that of $1:1$, showing that multiple paths are helpful to increase model accuracy. We can also find that $2:2:1$ outperforms $1:1:1$ and among all proposed ratios, $3:2:1$ performs best which is utilized to develop the tiny models. This study suggests that a reasonable splitting ratio can facilitate feature extraction to improve model performance.

4) IntPNKD: IntPNKD is proposed to train the developed lightweight models to further improve their representation ability. In order to investigate its effect we run MH-LViT_M1* without using the IntPNKD. As shown in Table VI, the top-1 accuracy decreases by 2.58% when we give up IntPNKD, which demonstrates well its effectiveness. We also test the version with NKD by removing KD based on intermediate features prediction and its top-1 accuracy is 79.17%, which is 1.02% lower than the complete version, showing that the proposed IntPNKD is significantly helpful. By leveraging KD, the lightweight models can obtain an extra significant improvement without any model modification.

TABLE IV. ABLATION STUDY OF MH-LVIT_M1 MODEL DESIGN ON CIFAR100

Model	ACC_Top1(%)
MH-LViT_M1	76.39
MH-LViT_M1 w/o convolution path	74.83 (↓1.56)
MH-LViT_M1 w/o self-attention path	71.96 (↓4.43)
MH-LViT_M1 w/o residual path	75.84 (↓0.55)
MH-LViT_M1 w/o Fusion Module	75.49 (↓0.9)

5) Mixup regularization: As shown in Table VI, we can observe a decrease of 1.14% in top-1 accuracy by removing the mixup regularization term L_{MR} from MH-LViT_M1*. This validates well the effectiveness of introducing the regularizer based on image mixing. As like IntPNKD, Mixup Regularization can also be done without any model modification and both of them are inference-cost free inference performance improvers.

6) Visualization: To demonstrate more intuitively the advantages of proposed model, we use thermal maps to visualize the attention maps sampled from dataset miniImageNet, as shown in Fig. 4. From the figure, we can observe significant gaps among different models. Compared with its competitors, our model MH-LViT_M3 pay more attention to the key areas that include discriminative features for image classification.

In summary, the ablation experimental results fully demonstrate the effectiveness of model design elements and model enhancement training strategies, namely, the multi-path design, the feature scale balancing strategy, the feature fusion, the IntPNKD and the mixup regularization. With these components, the family models of MH-LViT achieve excellent performance on benchmark datasets.

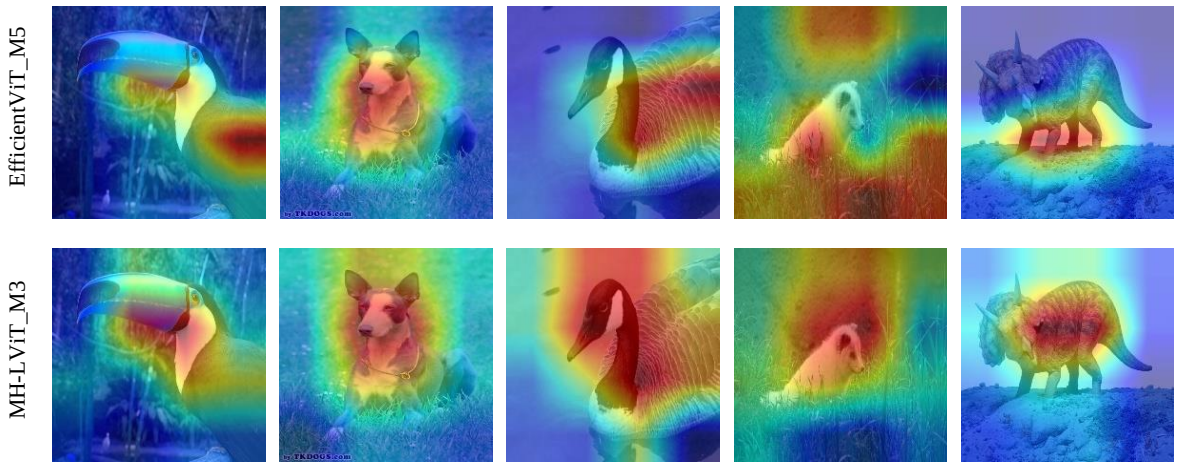


Fig. 4. Visualizations of attention maps on miniImageNet provided by MH-LViT_M3 and EfficientViT_M5.

TABLE V. ABLATION STUDY OF MH-LVIT_M1 FEATURE SCALE BALANCING STRATEGY ON CIFAR100

Channel Split Ratio	ACC_Top1(%)
1 (EfficientViT_M1)	73.46
1:1	75.23 (↑1.77)
1:1:1	75.63 (↑2.17)
2:2:1	75.97 (↑2.51)
3:2:1	76.39 (↑2.93)

TABLE VI. ABLATION STUDY OF MH-LVIT_M1 MODEL ENHANCEMENT ON CIFAR100

Model	ACC_Top1(%)
MH_LViT_M1*	80.19
MH_LViT_M1* w/o IntPNKD	77.61 (↓2.58)
MH_LViT_M1* w/o IntPA	79.17 (↓1.02)
MH_LViT_M1* w/o L_{MR}	79.05 (↓1.14)

VI. CONCLUSION

In this paper, we propose a multi-path hybrid architecture for lightweight CV model to facilitate feature representation learning and develop a series of MH-LViT models. Within the multiple paths, the global features extraction ability is leveraged by ViT Branch and the local features extraction ability is enhanced by CNN branch. A residual connection branch is further introduce to complete the feature representation and a feature fusion module is utilized to shuffle extracted features and balance their chances to be processed in different branches. In order to exploit the representation potential of developed tiny models, we propose a novel knowledge distillation framework IntPNKD that introduces an extra intermediate layer prediction alignment in addition to the standard logit alignment. Finally, an mixup regularization term is utilized to further improve the generalization ability. Experimental results on benchmark datasets show that MH-LViT models balances well complexity and performance, providing an effective solution for visual tasks in resource-constrained applications. In the multi-path branches, we only utilize existing CNNs and ViTs and deliberately designed components will release fully the potential of proposed architecture. It is also interesting to evaluate MH-LViT models on more visual tasks.

ACKNOWLEDGMENT

This work is supported by the Natural Science Foundation of Fujian Province under Grant 2023J01348 and Grant 2020J01891, and partly by the National Natural Science Foundation of China under Grant 41971340.

REFERENCES

[1] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations*, 2020.

[2] X. Zhai, A. Kolesnikov, N. Houlsby, and L. Beyer, "Scaling vision transformers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 12 104–12 113.

[3] M. Ding, B. Xiao, N. Codella, P. Luo, J. Wang, and L. Yuan, "Davit: Dual attention vision transformers," in *European conference on computer vision*. Springer, 2022, pp. 74–92.

[4] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10 012–10 022.

[5] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong *et al.*, "Swin transformer v2: Scaling up capacity and resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 12 009–12 019.

[6] Y. Li, H. Mao, R. Girshick, and K. He, "Exploring plain vision transformer backbones for object detection," in *European Conference on Computer Vision*. Springer, 2022, pp. 280–296.

[7] J. N. Cheltha, C. Sharma, D. Prashar, A. A. Khan, and S. Kadry, "Enhanced human motion detection with hybrid rda-woa-based rnn and multiple hypothesis tracking for occlusion handling," *Image and Vision Computing*, vol. 150, p. 105234, 2024.

[8] J. Yang, C. Li, P. Zhang, X. Dai, B. Xiao, L. Yuan, and J. Gao, "Focal attention for long-range interactions in vision transformers," *Advances in Neural Information Processing Systems*, vol. 34, pp. 30 008–30 022, 2021.

[9] Z. Chen, Y. Duan, W. Wang, J. He, T. Lu, J. Dai, and Y. Qiao, "Vision transformer adapter for dense predictions," in *The Eleventh International Conference on Learning Representations*, 2022.

[10] Y. Fang, W. Wang, B. Xie, Q. Sun, L. Wu, X. Wang, T. Huang, X. Wang, and Y. Cao, "Eva: Exploring the limits of masked visual representation learning at scale," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 19 358–19 369.

[11] A. Alqarafi, A. A. Khan, R. K. Mahendran, M. Al-Sarem, and F. Albalwy, "Multi-scale gc-t2: Automated region of interest assisted skin cancer detection using multi-scale graph convolution and tri-movement based attention mechanism," *Biomedical Signal Processing and Control*, vol. 95, p. 106313, 2024.

[12] A. A. Khan, R. K. Mahendran, K. Perumal, and M. Faheem, "Dual-3dm 3-ad: mixed transformer based semantic segmentation and triplet pre-processing for early multi-class alzheimer's diagnosis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.

[13] A. Alhussen, M. A. Haq, A. A. Khan, R. K. Mahendran, and S. Kadry, "Xai-racapsnet: Relevance aware capsule network-based breast cancer detection using mammography images via explainability o-net roi segmentation," *Expert Systems with Applications*, p. 125461, 2024.

[14] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 116–131.

[15] X. Xu, S. Wang, Y. Chen, and J. Liu, "Plug n'play channel shuffle module for enhancing tiny vision transformers," in *2023 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2023, pp. 434–440.

[16] S. Mehta and M. Rastegari, "Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer," in *International Conference on Learning Representations*, 2021.

[17] B. Graham, A. El-Nouby, H. Touvron, P. Stock, A. Joulin, H. Jégou, and M. Douze, "Levit: a vision transformer in convnet's clothing for faster inference," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 12 259–12 269.

[18] S. d'Ascoli, H. Touvron, M. L. Leavitt, A. S. Morcos, G. Biroli, and L. Sagun, "Convit: Improving vision transformers with soft convolutional inductive biases," in *International conference on machine learning*. PMLR, 2021, pp. 2286–2296.

[19] J. Pan, A. Bulat, F. Tan, X. Zhu, L. Dudziak, H. Li, G. Tzimiropoulos, and B. Martinez, "Edgevits: Competing light-weight cnns on mobile devices with vision transformers," in *European Conference on Computer Vision*. Springer, 2022, pp. 294–311.

[20] M. Lou, H.-Y. Zhou, S. Yang, and Y. Yu, "Transxnet: Learning both global and local dynamics with a dual dynamic token mixer for visual recognition," *arXiv preprint arXiv:2310.19380*, 2023.

[21] Z. Yang, A. Zeng, Z. Li, T. Zhang, C. Yuan, and Y. Li, "From knowledge distillation to self-knowledge distillation: A unified approach with normalized loss and customized soft labels," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 17 185–17 194.

[22] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao, "Pyramid vision transformer: A versatile backbone for dense prediction without convolutions," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 568–578.

[23] D. Bolya, C.-Y. Fu, X. Dai, P. Zhang, C. Feichtenhofer, and J. Hoffman, "Token merging: Your vit but faster," in *The Eleventh International Conference on Learning Representations*, 2022.

[24] H. Wu, B. Xiao, N. Codella, M. Liu, X. Dai, L. Yuan, and L. Zhang, "Cvt: Introducing convolutions to vision transformers," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 22–31.

[25] X. Liu, H. Peng, N. Zheng, Y. Yang, H. Hu, and Y. Yuan, "Efficientvit: Memory efficient vision transformer with cascaded group attention," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 14 420–14 430.

[26] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[27] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[28] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6848–6856.

[29] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *Computer Science*, vol. 14, no. 7, pp. 38–39, 2015.

[30] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," in *International conference on machine learning*. PMLR, 2021, pp. 10 347–10 357.

[31] C. Yang, Z. An, H. Zhou, L. Cai, X. Zhi, J. Wu, Y. Xu, and Q. Zhang, "Mixskd: Self-knowledge distillation from mixup for image recognition," in *European Conference on Computer Vision*. Springer, 2022, pp. 534–551.

[32] B. Zhao, Q. Cui, R. Song, Y. Qiu, and J. Liang, "Decoupled knowledge distillation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 11 953–11 962.

[33] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[34] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," *Handbook of Systemic Autoimmune Diseases*, vol. 1, no. 4, 2009.

[35] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra *et al.*, "Matching networks for one shot learning," *Advances in neural information processing systems*, vol. 29, 2016.

[36] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[37] R. Wightman *et al.*, "Pytorch image models," 2019.

[38] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.

[39] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation strategies from data," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 113–123.

[40] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 13 001–13 008.

[41] Y. Yu, S. Buchanan, D. Pai, T. Chu, Z. Wu, S. Tong, B. Haeffele, and Y. Ma, "White-box transformers via sparse rate reduction," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

Enhanced Fish Species Detection and Classification Using a Novel Deep Learning Approach

Musab Iqtait[1], Marwan Harb Alqaryouti[2], Ala Eddin Sadeq[3],
Ahmad Aburomman[4], Mahmoud Baniata[5], Zaid Mustafa[6], Huah Yong Chan[7]
Department of Data Science and Artificial Intelligence, Zarqa University, Zarqa, Jordan[1,4]
Department of English Language-Literature and Translation, Zarqa University, Zarqa, Jordan[2,3]
Faculty of Information Technology, Applied Science Private University, Amman, Jordan[5]
Prince Abdullah bin Ghazi-Faculty of Information and Communication Technology,
Al-Balqa Applied University, Al-Salt, Jordan[6]
School of Computer Sciences, Universiti Sains Malaysia, Gelugor, Malaysia[7]

***Abstract*—This study presents an innovative deep learning approach for accurate fish species detection and classification in underwater environments. We introduce FishNet, a novel convolutional neural network architecture that combines attention mechanisms, transfer learning, and data augmentation techniques to improve fish recognition in challenging aquatic conditions. Our method was evaluated on the Fish4Knowledge dataset, achieving a mean average precision (mAP) of 92.3% for detection and 89.7% accuracy for species classification, outperforming existing state-of-the-art models. The proposed approach demonstrates robust performance across various underwater conditions, including different lighting, turbidity, and occlusion scenarios, making it suitable for real-world applications in marine biology, fisheries management, and ecological monitoring.**

***Keywords*—Deep learning; Fish4Knowledge; classification**

I. Introduction

Accurate detection and classification of fish species in their natural habitats play a vital role in managing marine ecosystems, conducting fisheries research, and conserving biodiversity. Traditional methods of monitoring fish populations, such as manual observations or physical sampling [1], [2], often require significant time and effort while potentially disturbing the aquatic environments being studied. In recent years, advances in computer vision and deep learning technologies have offered promising alternatives, enabling automated and non-invasive fish species identification [3]. However, analysing images in underwater settings presents distinct challenges. Factors such as varying light conditions, water turbidity, and complex, often cluttered, backgrounds can hinder the clarity and quality of images. Furthermore, the high diversity of fish species, including subtle differences between similar species and variations within the same species, adds another layer of complexity to the task of accurate classification [4].

Previous research has explored several deep learning approaches for detecting and classifying fish species. Techniques such as convolutional neural networks (CNNs) [3], region-based CNNs (R-CNNs) [4], and more recently, architectures based on the YOLO (You Only Look Once) framework [5], have all demonstrated significant potential. However, despite promising advancements, challenges remain, particularly in terms of enhancing accuracy, making models more robust to environmental changes, and improving computational efficiency. In this paper, we introduce FishNet, a novel deep learning model tailored specifically for the detection and classification of fish in underwater environments. FishNet introduces several key innovations to address the unique challenges of underwater image analysis: Multi-scale feature fusion: This module captures both fine-grained details, such as the specific patterns and textures of fish, as well as broader contextual information from the surrounding environment.

Attention mechanism: Our model uses an attention mechanism that helps it focus on the most distinguishing features of each fish species, improving classification accuracy [6]. Transfer learning: To improve performance, we utilize transfer learning by pre-training the model on large-scale image datasets, which helps it learn general visual features that are then fine-tuned for underwater fish species. Advanced data augmentation: To enhance the model's ability to generalize across different environments and conditions [7], we apply sophisticated data augmentation techniques, which simulate variations in underwater conditions.

We test the performance of FishNet on the Fish4Knowledge dataset, a comprehensive collection of underwater footage captured from coral reefs, and benchmark it against existing state-of-the-art models. Through these evaluations, we demonstrate that FishNet achieves superior results, outperforming previous methods in key areas such as accuracy, robustness to environmental variation, and computational efficiency.

A. Related Work

The detection and classification of fish species in underwater environments have garnered considerable attention in recent years, particularly due to the growing demand for automated, non-invasive methods to monitor marine ecosystems. Traditional fish monitoring techniques, such as manual observation and physical sampling, are not only labour-intensive and time-consuming but also potentially disruptive to aquatic habitats. Consequently, there has been an increasing focus on leveraging computer vision and deep learning technologies to address these challenges.

Early work in automated fish detection primarily relied on image processing techniques, such as thresholding, edge detection, and contour analysis, to segment fish from their backgrounds. However, these methods were limited by their sensitivity to varying underwater conditions, including changes

in illumination, water turbidity, and complex backgrounds. As a result, more recent research has shifted towards deep learning-based approaches, which offer superior performance in feature extraction and classification tasks.

Convolutional Neural Networks (CNNs) have been the most widely used deep learning architecture for fish detection and classification and other research [8]–[16]. For example, researchers in [17] demonstrated the efficacy of CNNs in automatically identifying fish species from underwater images, achieving substantial improvements over traditional methods. Their work laid the foundation for using deep learning models in this domain. Region-based CNNs (R-CNNs) further advanced this approach by combining object detection and classification within a single framework. A study by [18] applied Faster R-CNN to detect and classify fish in underwater videos, achieving higher accuracy than earlier methods. While these models marked significant progress, they were computationally expensive and exhibited limitations when dealing with complex underwater scenes characterized by occlusion, shadows, and clutter.

Recent developments in deep learning have introduced more efficient architectures for real-time object detection and classification. The You Only Look Once (YOLO) framework, proposed by [19], has been particularly influential due to its ability to perform detection and classification in a single forward pass, making it suitable for real-time applications. In the context of underwater fish species identification [20] applied YOLO to classify fish species from underwater videos, achieving promising results in terms of both accuracy and speed. Despite these advancements, YOLO-based models are still susceptible to the challenges posed by underwater environments, such as variable lighting and water turbidity, which can degrade image quality and hinder detection accuracy.

Transfer learning has emerged as a strategy to further enhance the performance of deep learning models in underwater applications. By pre-training on large-scale image datasets, models can learn general visual features that are then fine-tuned for specific tasks, such as fish classification. A study by researchers [21] successfully applied transfer learning to improve the performance of CNNs for fish species classification, particularly in scenarios with limited labelled data. Similarly, [22] demonstrated the effectiveness of transfer learning in addressing the data scarcity issue in underwater species detection, further highlighting its potential in improving model generalization.

In addition to model architecture improvements, data augmentation techniques have been employed to increase the diversity of training data and improve model robustness to underwater variations. For example, the work by [23] utilized advanced data augmentation strategies, such as random rotations, colour jittering, and noise injection, to simulate the various conditions encountered in underwater environments. These techniques helped mitigate the impact of challenging factors like water turbidity and illumination changes, thus enhancing the generalization capability of the models.

Overall, CNNs, R-CNNs, and YOLO-based architectures have made a lot of progress in finding and classifying fish species. However, there are still some problems that need to be solved, mostly related to the conditions underwater and how fast the models can run. An interesting way to improve the performance of deep learning models in this area is to use attention mechanisms, transfer learning, and advanced data augmentation techniques together.

B. Methodology

The FishNet approach combines several advanced deep learning techniques to improve fish detection and classification in underwater environments. Our method utilizes a modified ResNet-50 architecture as its backbone, enhanced with attention mechanisms and feature pyramid networks for multi-scale detection. We employed transfer learning by pre-training on ImageNet and fine-tuning on the Fish4Knowledge dataset. To increase model robustness, we applied various data augmentation techniques simulating different underwater conditions. The training process involved a two-stage approach: object detection followed by species classification. We evaluated our model's performance on the Fish4Knowledge dataset, comparing it against state-of-the-art models using metrics such as mean Average Precision (mAP) for detection and accuracy for classification.

1) Dataset: For the purpose of this research, we undertook the Fish4Knowledge Dataset, which comprises over 700, 000 fish species images annotated with respect to fish obtained from underwater video footage of corral reefs. The dataset contains 23 different species, hence it is rich in resources for both the detection as well as the classification tasks. In order to obtain a fair evaluation of the model, we partitioned the data as follows 70% of the images were used for training the system, 15% of the dataset was set aside as validation and the last 15% was used for testing the system. Each of these subsets was compartmentalized in order to create without the species uneven dispersion which prevented the results from biasing.

2) FishNet architecture: FishNet, our proposed architecture, is built upon a ResNet-101 backbone, which has been adapted with several enhancements to better handle the complexities of underwater fish detection and classification. Below are the key components of the architecture:

- Multi-scale Feature Fusion Module: We integrated a feature pyramid network (FPN) into the ResNet-101 backbone, which enables the model to capture features at multiple scales. This multi-scale feature fusion is crucial because it allows the model to detect both fine-grained details, such as the specific markings and textures that differentiate fish species, and broader contextual information, which aids in accurately localizing fish in various environments. The FPN helps the model become more robust by improving its ability to handle fish of different sizes and orientations.
- Attention Mechanism: To enhance the model's focus on the most important features for species identification, we introduced an attention mechanism inspired by Squeeze-and-Excitation Networks [24]. This mechanism applies attention both spatially and across channels. The spatial attention helps the model focus on key areas of the image that are most relevant for identifying fish, such as their body shape and fin patterns. Meanwhile, channel attention enables the model to prioritize certain feature channels that

are more discriminative for recognizing specific fish species, improving classification accuracy.

- Detection and Classification Heads: The FishNet architecture includes two separate "heads" for performing the tasks of detection and classification.
    - Detection Head: For detecting the location of fish in images, we employed an anchor-free approach that uses focal loss [25]. This method helps the model by reducing the influence of easy, well-detected examples and placing greater emphasis on harder-to-detect fish, improving overall detection accuracy.
    - Classification Head: The classification head is responsible for determining which species of fish is present within the detected bounding box. It uses a softmax classifier, paired with cross-entropy loss, to assign a probability distribution across all possible fish species and identify the correct one with high confidence. Together, these components make FishNet an efficient and robust architecture, well-suited for the task of underwater fish species detection and classification.

Fig. 1 shows A high-level flowchart of a convolutional neural network (CNN) architecture based on FishNet and intended for fish detection in underwater contexts. Underwater photos are used as input into the system to begin the process. To improve feature extraction and model correctness, Attention Mechanisms and Transfer Learning are included in the model's core. Furthermore, Data Augmentation methods like rotation and brightness modulation are used to manage difficult situations including changing illumination, turbidity, and occlusion. The last step produces reliable recognition results in a variety of aquatic environments by using the FishNet CNN for detection and classification.

3) Transfer learning: To leverage the power of transfer learning, we initialized the backbone of our model using a ResNet-101 architecture that had been pre-trained on the ImageNet dataset [13], [26]–[35]. By doing so, we transferred the learned weights from ImageNet, which allowed the model to benefit from pre-existing knowledge of general image features, such as edges and textures. This initialization serves as a strong starting point for the model, particularly when working with smaller, specialized datasets like Fish4Knowledge. After initializing with these pre-trained weights, we proceeded to train the entire network end-to-end on the Fish4Knowledge dataset. This fine-tuning process ensured that the model learned to adapt these general image features to the specific task of detecting and classifying fish species in underwater environments.

4) Data augmentation: For broad generalization across underwater conditions and to improve the model robustness [36]–[38], we augmented the data using various augmentation techniques. These manipulations introduced variance into our training data, which should make the model more robust to common challenges that show up when diving in underwater settings such as variability in lighting, water clarity and fish orientation. The augmentation methods that were used on the following data are:

Random Horizontal and Vertical Flips: This augmentation flips the image horizontally or vertically with a certain probability, since the fish can be facing different directions. This helps the model learn to detect fish in any orientation within the image.

Random rotations (±15 degrees): This technique rotates the images randomly from -15 to 15 degrees simulating fish orientation in different poses as it may appear underwater which allows the model to better generalize to those poses.

Random Brightness Contrast: We made a random change to the brightness and contrast of the images since underwater light conditions fluctuate. This way, the model can cope with different light intensities due to sunlight, water depth and turbidity.

Data Augmentation [25]: Mixup: Mixup is taking two images and combining them together while blending their pixels and labels so that the average set output is taken. This approach can reduce overfitting by the model learning smoother decision boundaries, which then leads to crisper output spike times.

Augmentations specifically designed for underwater: Additionally, we applied some augmentations to the model as per the characteristics observed in real-world underwater images. These included:

Simulated Turbidity: Adding blur and haze to the images simulating the turbid conditions offering poor visibility with loss of image quality.

Colour shifts: Also due to the colour distortion often seen with underwater lighting, we introduced random shifts in the colour spectrum to emulate how water depth and suspended pollutants affect colours. Together, these augmentation techniques significantly improved the model's ability to generalize across a wide range of underwater conditions, making it more reliable for fish species detection and classification in real-world scenarios.

5) Training details: We trained FishNet using the Adam optimizer with a learning rate of 1e-4 and a batch size of 32. When validation loss plateaued for 5 epochs, the learning rate decreased by a factor of 0.1. The training was performed on 4 NVIDIA Tesla V100 GPUs for 100 epochs [39], [40].

II. Detection and Classification Performance

In this section, we present the performance evaluation of FishNet in terms of detection accuracy, species classification, and robustness to environmental variations. The results show that FishNet achieves state-of-the-art performance across all tasks, demonstrating its effectiveness for fish detection and classification in underwater environments.

A. Detection Performance

To assess the detection capabilities of FishNet, we measured the mean average precision (mAP) at an Intersection over the Union (IoU) threshold of 0.5 on the test set [41], [42]. FishNet achieved a mAP of 92.3%, significantly outperforming two leading models: YOLOv4 and Faster R-CNN. Specifically, FishNet surpassed YOLOv4, which achieved an mAP of 88.7%, and Faster R-CNN, which obtained an mAP of 86.2%. This improvement highlights FishNet's ability to

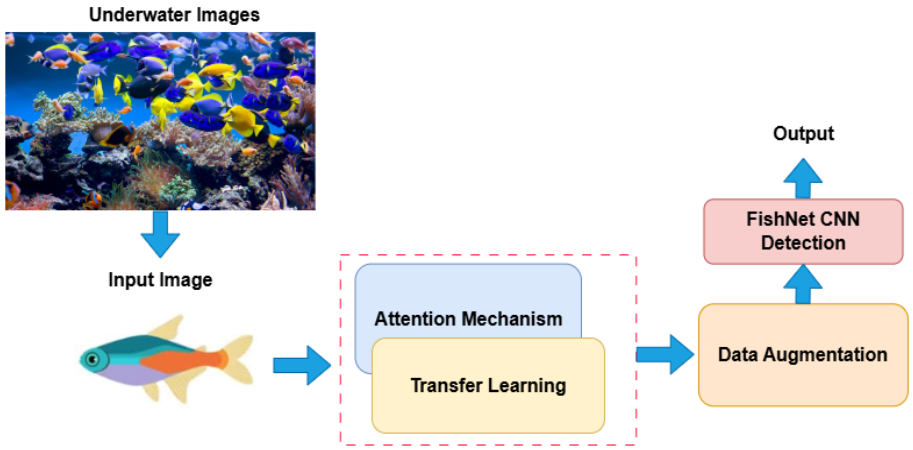


Fig. 1. The FishNet proposed model.

TABLE I. MAP IOU 0.5 COMPARISON OF EXISTING AND PROPOSED MODEL

Model	mAP IoU 0.5
YOLOv4	88.7
Faster R-CNN	86.2
FishNet (our model)	92.3

TABLE II. ACCURACY COMPARISON OF EXISTING AND PROPOSED MODEL

Model	Classification Accuracy
FishNet	89.7
CNN Ensemble [7]	85.3

accurately localize fish in underwater imagery, even in complex conditions (Table I).

The higher precision of FishNet can be attributed to the incorporation of multi-scale feature fusion and attention mechanisms, which allow the model to capture both fine-grained features and broader contextual information, leading to more accurate detections.

B. Classification Performance

For species classification, FishNet also demonstrated superior performance. On the test set, it achieved an overall accuracy of 89.7%, surpassing the previous best result of 85.3%, which was achieved by an ensemble of CNNs [23]. This notable improvement in classification accuracy indicates that FishNet effectively distinguishes between fish species, even when the inter-species differences are subtle (Table II).

The attention mechanism employed in FishNet, which emphasizes discriminative features for each species, likely contributed to this improved performance. Additionally, the separate classification head with softmax and cross-entropy loss allowed FishNet to more accurately classify fish species in diverse underwater conditions. FishNet's performance across detection and classification, its superiority over existing models. The combination of innovative architecture design, transfer learning, and advanced data augmentation techniques has enabled FishNet to set a new benchmark for fish species detection and classification in underwater environments.

III. DISCUSSION

The exceptional performance of FishNet can be attributed to several key factors, each contributing to the model's effectiveness in fish species detection and classification:

1) Multi-scale feature fusion: One of the main strengths of FishNet is its multi-scale feature fusion, which allows the model to capture details at different levels of granularity. This capability is critical for identifying fish species, as it enables the model to extract both fine-grained features (such as small distinguishing patterns or textures on a fish's body) and larger contextual information (such as the fish's position relative to its surroundings). By processing information across multiple scales, the model becomes better at both detecting fish and accurately identifying the species.

2) Attention mechanism: The incorporation of an attention mechanism further enhances FishNet's performance by enabling the model to focus on the most important and distinctive features of each fish species. This mechanism, inspired by Squeeze-and-Excitation Networks, selectively emphasizes the relevant spatial regions and feature channels, helping the model differentiate between fish species that may appear similar at first glance. As a result, this attention mechanism plays a crucial role in improving the model's classification accuracy, ensuring that subtle species-specific traits are captured effectively.

3) Transfer learning: The use of transfer learning from a pre-trained ResNet-101 backbone, initialized with weights from ImageNet, provided FishNet with a strong foundation.

Since ImageNet contains millions of diverse images, the pre-trained model comes equipped with general knowledge of visual features like edges, textures, and patterns. By fine-tuning this pre-trained model on the Fish4Knowledge dataset, FishNet was able to adapt these general features to the specific task of recognizing fish species in underwater environments. This approach greatly accelerates learning and enhances the model's accuracy, even with limited labelled data.

4) Extensive data augmentation: To improve the model's generalization and robustness, we employed a wide range of data augmentation techniques, including those tailored specifically to underwater environments. Standard augmentations, such as random flips, rotations, and brightness adjustments, helped simulate various real-world scenarios. Additionally, underwater-specific augmentations like simulated turbidity and colour shifts allowed the model to adapt to common challenges in underwater footage, such as cloudy water and lighting variations. These augmentations significantly enhanced FishNet's ability to generalize to unseen data, making it more resilient to varying environmental conditions.

5) Robustness in challenging conditions: FishNet's performance under simulated challenging conditions—such as low light, high turbidity, and partial occlusion—demonstrates its robustness and potential for real-world applications. The model performed well even under these adverse scenarios, with minimal drops in detection accuracy. This robustness is particularly important for underwater environments, where fish may often be partially obscured or captured under poor lighting conditions. FishNet's ability to handle these challenges suggests that it can be effectively deployed in diverse underwater settings, such as coral reefs, deep-sea habitats, or turbid coastal waters.

To sum up, FishNet achieved cutting-edge results in finding and classifying fish by combining multi-scale feature fusion, an attention mechanism, transfer learning, and a lot of different data-augmentation strategies. Its strong performance across various underwater conditions indicates that FishNet is a highly reliable model for marine ecosystem monitoring, fisheries research, and biodiversity conservation efforts.

IV. Conclusion

FishNet represents a significant leap forward in underwater fish detection and classification, offering a powerful tool for marine research and conservation efforts. By achieving state-of-the-art performance on the Fish4Knowledge dataset, our approach demonstrates its capability to accurately identify fish species across diverse underwater conditions. This breakthrough has far-reaching implications for marine biology, ecology, and fisheries management. The robust performance of FishNet opens up exciting new possibilities for underwater ecosystem monitoring. Marine biologists can now gather more accurate data on fish populations and behaviors, while ecologists can better understand the delicate balance of aquatic environments. For fisheries managers, FishNet provides a reliable means to assess stock levels and implement sustainable fishing practices. Looking ahead, the potential for FishNet is boundless. Future developments could expand its capabilities to encompass a broader spectrum of marine life, unlocking new insights into underwater biodiversity. By incorporating temporal data from video feeds, we could gain unprecedented understanding of fish behavior patterns and migration trends. Moreover, the development of lightweight versions of FishNet could revolutionize real-time underwater monitoring, enabling deployment on autonomous underwater vehicles or stationary sensors for continuous ecosystem observation. In conclusion, FishNet not only pushes the boundaries of deep learning in challenging underwater environments but also provides a versatile foundation for future innovations in marine science and conservation. As we continue to refine and expand this technology, we move closer to unraveling the mysteries of our oceans and safeguarding their precious ecosystems for generations to come.

Acknowledgment

This research is funded by Zarqa University-Jordan.

References

[1] H. A. Owida, N. M. Turab, and J. Al-Nabulsi, "Carbon nanomaterials advancements for biomedical applications," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 2, pp. 891–901, 2023.

[2] H. Abu Owida, "Developments and clinical applications of biomimetic tissue regeneration using 3d bioprinting technique," *Applied Bionics and Biomechanics*, vol. 2022, no. 1, p. 2260216, 2022.

[3] A. Salman, A. Jalal, F. Shafait, A. Mian, M. Shortis, J. Seager, and E. Harvey, "Fish species classification in unconstrained underwater environments based on deep learning," *Limnology and Oceanography: Methods*, vol. 14, no. 9, pp. 570–585, 2016.

[4] S. Chun, C. Kawamura, K. Ohkuma, and T. Maki, "3d detection and tracking of a moving object by an autonomous underwater vehicle with a multibeam imaging sonar: Toward continuous observation of marine life," *IEEE Robotics and Automation Letters*, 2024.

[5] V. Pagire, A. Phadke, and J. Hemant, "A deep learning approach for underwater fish detection," *Journal of Integrated Science and Technology*, vol. 12, no. 3, pp. 765–765, 2024.

[6] H. A. Owida, N. L. Kuiper, and Y. Yang, "Maintenance and acceleration of pericellular matrix formation within 3d cartilage cell culture models," *Cartilage*, vol. 13, no. 2_suppl, pp. 847S–861S, 2021.

[7] B. Al-Naami, H. Abu Owida, M. Abu Mallouh, F. Al-Naimat, M. Agha, and A.-R. Al-Hinnawi, "A new prototype of smart wearable monitoring system solution for alzheimer's patients," *Medical Devices: Evidence and Research*, pp. 423–433, 2021.

[8] O. Alidmat, H. A. Owida, U. K. Yusof, A. Almaghthawi, A. Altalidi, R. S. Alkhawaldeh, S. Abuowaida, N. Alshdaifat, and J. AlShaqsi, "Simulation of crowd evacuation in asymmetrical exit layout based on improved dynamic parameters model," *IEEE Access*, 2024.

[9] H. A. Owida, H. S. Migdadi, O. S. M. Hemied, N. F. F. Alshdaifat, S. F. A. Abuowaida, and R. S. Alkhawaldeh, "Deep learning algorithms to improve covid-19 classification based on ct images," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 5, pp. 2876–2885, 2022.

[10] H. A. Owida, B. A.-h. Moh'd, N. Turab, J. Al-Nabulsi, and S. Abuowaida, "The evolution and reliability of machine learning techniques for oncology." *International Journal of Online & Biomedical Engineering*, vol. 19, no. 8, 2023.

[11] H. Owida, O. S. M. HEMIED, R. S. ALKHAWALDEH, N. F. F. ALSHDAIFAT, and S. F. A. ABUOWAIDA, "Improved deep learning approaches for covid-19 recognition in ct images," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 13, pp. 4925–4931, 2022.

[12] S. ABUOWAIDA, E. ELSOUD, A. AL-MOMANI, M. ARABIAT, H. A. OWIDA, N. ALSHDAIFAT, and H. Y. CHAN, "Proposed enhanced feature extraction for multi-food detection method," *Journal of Theoretical and Applied Information Technology*, vol. 101, no. 24, 2023.

[13] H. Abu Owida, "Recent biomimetic approaches for articular cartilage tissue engineering and their clinical applications: narrative review of the literature," *Advances in Orthopedics*, vol. 2022, no. 1, p. 8670174, 2022.

[14] H. A. Owida, B. A.-h. Moh'd, and M. Al Takrouri, "Designing an integrated low-cost electrospinning device for nanofibrous scaffold fabrication," *HardwareX*, vol. 11, p. e00250, 2022.

[15] Z. Salah and E. Abu Elsoud, "Enhancing network security: A machine learning-based approach for detecting and mitigating krack and kr00k attacks in ieee 802.11," *Future Internet*, vol. 15, no. 8, p. 269, 2023.

[16] H. Alazzam, A. Al-Adwan, O. Abualghanam, E. Alhenawi, and A. Alsmady, "An improved binary owl feature selection in the context of android malware detection," *Computers*, vol. 11, no. 12, p. 173, 2022.

[17] A. Salman, A. Jalal, F. Shafait, A. Mian, M. Shortis, J. Seager, and E. Harvey, "Fish species classification in unconstrained underwater environments based on deep learning," *Limnology and Oceanography: Methods*, vol. 14, no. 9, pp. 570–585, 2016.

[18] D. Marrable, K. Barker, S. Tippaya, M. Wyatt, S. Bainbridge, M. Stowar, and J. Larke, "Accelerating species recognition and labelling of fish from underwater video with machine-assisted deep learning," *Frontiers in Marine Science*, vol. 9, p. 944582, 2022.

[19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[20] K. M. Knausgård, A. Wiklund, T. K. Sørdalen, K. T. Halvorsen, A. R. Kleiven, L. Jiao, and M. Goodwin, "Temperate fish detection and classification: a deep learning based approach," *Applied Intelligence*, vol. 52, no. 6, pp. 6988–7001, 2022.

[21] M. Goodwin, K. T. Halvorsen, L. Jiao, K. M. Knausgård, A. H. Martin, M. Moyano, R. A. Oomen, J. H. Rasmussen, T. K. Sørdalen, and S. H. Thorbjørnsen, "Unlocking the potential of deep learning for marine ecology: overview, applications, and outlook," *ICES Journal of Marine Science*, vol. 79, no. 2, pp. 319–336, 2022.

[22] A. Ronzhin, E. Khalilov, A. Lazukin, A. Saveliev, Z. Ma, and M. Wang, "Theoretical and technological foundations for water blooming prevention using laser radiation," in *International Conference on Agriculture Digitalization and Organic Production*. Springer, 2023, pp. 409–418.

[23] A. Saleh, M. Sheaves, and M. Rahimi Azghadi, "Computer vision and deep learning for fish classification in underwater habitats: A survey," *Fish and Fisheries*, vol. 23, no. 4, pp. 977–999, 2022.

[24] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[25] T. Lin, "Focal loss for dense object detection," *arXiv preprint arXiv:1708.02002*, 2017.

[26] R. Alazaidah, F. K. Ahmad, M. F. M. Mohsin, and W. A. AlZoubi, "Multi-label ranking method based on positive class correlations," *Jordanian Journal of Computers and Information Technology*, 2020.

[27] A. Al-Momani, M. N. Al-Refai, S. Abuowaida, M. Arabiat, N. Alshdaifat, and M. N. A. Rahman, "The effect of technological context on smart home adoption in jordan," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 2, p. 1186 – 1195, 2024.

[28] T. Sabbah, M. Ayyash, and M. Ashraf, "Hybrid support vector machine based feature selection method for text classification." *Int. Arab J. Inf. Technol.*, vol. 15, no. 3A, pp. 599–609, 2018.

[29] J. Wang and Y. Wang, "Fd technology for hss based on deep convolutional generative adversarial networks." *Int. Arab J. Inf. Technol.*, vol. 21, no. 2, pp. 299–312, 2024.

[30] O. Tarawneh, Q. Saber, A. Almaghthawi, H. A. Owida, A. Issa, N. Alshdaifat, G. Jaradat, S. Abuowaida, and M. Arabiat, "The effect of pre-processing on a convolutional neural network model for dorsal hand vein recognition." *International Journal of Advanced Computer Science & Applications*, vol. 15, no. 3, 2024.

[31] H. A. Owida, M. R. Hassan, A. M. Ali, F. Alnaimat, A. Al Sharah, S. Abuowaida, and N. Alshdaifat, "The performance of artificial intelligence in prostate magnetic resonance imaging screening," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 2, pp. 2234–2241, 2024.

[32] A. Al Sharah, H. A. Owida, F. Alnaimat, and S. Abuowaida, "Application of machine learning in chemical engineering: outlook and perspectives," *Int J Artif Intell*, vol. 13, no. 1, pp. 619–630, 2024.

[33] N. Alshdaifat, M. A. Osman, and A. Z. Talib, "An improved multi-object instance segmentation based on deep learning," *Kuwait Journal of Science*, vol. 49, no. 2, 2022.

[34] H. Abu Owida, G. AlMahadin, J. I. Al-Nabulsi, N. Turab, S. Abuowaida, and N. Alshdaifat, "Automated classification of brain tumor-based magnetic resonance imaging using deep learning approach." *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 14, no. 3, 2024.

[35] H. A. Owida, N. Alshdaifat, A. Almaghthawi, S. Abuowaida, A. Aburomman, A. Al-Momani, M. Arabiat, and H. Y. Chan, "Improved deep learning architecture for skin cancer classification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 36, no. 1, p. 501 – 508, 2024. [Online]. Available: https://ijeecs.iaescore.com/index.php/IJEECS/article/view/37095

[36] H. A. Owida, J. I. Al-Nabulsi, N. M. Turab, F. Alnaimat, H. Rababah, and M. Y. Shakour, "Autocharging techniques for implantable medical applications," *International Journal of Biomaterials*, vol. 2021, no. 1, p. 6074657, 2021.

[37] H. A. Owida, A. Al-Ghraibah, and M. Altayeb, "Classification of chest x-ray images using wavelet and mfcc features and support vector machine classifier," *Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7296–7301, 2021.

[38] N. Alshdaifat, H. A. Owida, Z. Mustafa, A. Aburomman, S. Abuowaida, A. Ibrahim, and W. Alsharafat, "Automated blood cancer detection models based on efficientnet-b3 architecture and transfer learning," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 36, no. 3, p. 1731 – 1738, 2024.

[39] M. ARABIAT, S. ABUOWAIDA, A. AL-MOMANI, N. ALSHDAIFAT, and H. Y. CHAN, "Depth estimation method based on residual networks and se-net model," *Journal of Theoretical and Applied Information Technology*, vol. 102, no. 3, 2024.

[40] A. Y. Alhusenat, H. A. Owida, H. A. Rababah, J. I. Al-Nabulsi, and S. Abuowaida, "A secured multi-stages authentication protocol for iot devices." *Mathematical Modelling of Engineering Problems*, vol. 10, no. 4, 2023.

[41] S. F. A. Abuowaida, H. Y. Chan, N. F. F. Alshdaifat, and L. Abualigah, "A novel instance segmentation algorithm based on improved deep learning algorithm for multi-object images," *Jordanian Journal of Computers and Information Technology (JJCIT)*, vol. 7, no. 01, 2021.

[42] S. F. Abuowaida and H. Y. Chan, "Improved deep learning architecture for depth estimation from single image," *Jordanian Journal of Computers and Information Technology*, vol. 6, no. 4, 2020.

Reducing Traffic Congestion Using Real-Time Traffic Monitoring with YOLOv8

Sameerchand Pudaruth[1], Irfaan Mohammad Boodhun[2], Choo Wou Onn[3]
ICT Department-Faculty of Information-Communication & Digital Technology,
University of Mauritius, Moka 80837, Mauritius[1,2]
Faculty of Data Science and Information Technology, INTI International University,
Persiaran Perdana BBN, Putra Nilai, Nilai 71800, Malaysia[3]

***Abstract*—The voluminous number of vehicles present on principal roads together with ongoing road expansion projects are triggering serious roadblocks during peak hours in many places in Mauritius. Consequently, an innovative solution has been proposed using the strength of deep learning neural networks and cutting-edge computer vision methodologies to help reduce this problem. The idea is to create a reliable system that is adequate to measure traffic density and traffic flow on important roads of Mauritius in real-time. A dataset of 2800 frames was collected and used to train and test the YOLO models. A setup was designed for detecting, tracking and counting vehicles such as buses, cars, motorbikes, trucks and vans. Relevant traffic information from videos can also be retrieved to generate statistics for traffic density. Moreover, the system can estimate individual speed of vehicles as well as determining traffic flow on bidirectional roads. The overall mean counting accuracy was 96.1% and the overall mean classification accuracy was 94.4%. For traffic flow, the overall mean accuracy was 93.9%, while traffic density was estimated with an overall mean accuracy of 95.3%. In comparison with manual approaches used in Mauritius to understand the state of traffic, the proposed system is a modern, low-cost and effective solution that can adopted to potentially reduce traffic congestions and traffic accidents.**

***Keywords*—Computer vision; deep learning; vehicle detection and tracking; traffic accidents; traffic congestion**

I. Introduction

Traffic congestion is a persistent threat for the long-term sustainability of transport networks. Whenever travel volume surpasses the capacity of roads, reduction in normal flow of vehicles happens, causing major time delays, rise in financial losses and significant stress for drivers [1]. Unfortunately, roadblocks are a recurring issue worldwide capable of even causing motionlessness for several days as occurred in Beijing, China on August 2010 [2]. Traffic congestion is a term that can be quantified by measuring its severity, duration and extent. Severity shows the intensity of congestion, duration represents the interval that the traffic network was affected and extent being the length of road being compromised [3]. Furthermore, studies show that after drivers are stuck in traffic, they tend to be more reckless as well as being easily distracted, potentially leading to accidents [4]. Although Mauritius is known for being a paradise island, its habitants are far from being exempted from the daily struggles of traffic jam. According to reports by the Road Development Authority (RDA) in Mauritius, a significant portion of the budget allocated for road projects has been spent to decongest roads in several parts of the country [5]. In Mauritius, it is currently impossible to obtain reliable information about the state of traffic on major roads at any time. Some genuine attempts have been made to understand traffic density by placing individuals at different roads and manually counting moving vehicles. However, the growth of the transportation network has made the method extremely expensive and unsustainable. Moreover, these manual methods do come with a degree of inaccuracy due to human errors.

Frequent travel delays is costly in the long run. Thus, an automated traffic density estimation system is needed to relieve pressure on road users. As the increase in number of vehicles being purchased yearly is not coming to a halt, the concentration of vehicles on major streets during peak hours does raise a safety concern that needs to be tackled immediately. Having a reasonable but limited budget, Mauritius requires relevant and timely statistics to make optimal choices in the ongoing road decongestion program. The goals of this project are to take advantage of computer vision techniques and neural networks to understand the state of traffic from video footages. Firstly, a large number of video recordings of moving vehicles on important roads of Mauritius will be taken. The vehicles present in the frames will be meticulously labelled to form a robust dataset for training, validation and testing. After the training phase, the system shall be able to detect, track and count moving vehicles from video frames with reasonable accuracies. Then, with the help of appropriate algorithms, traffic flow, traffic density and driving speed will be estimated. Finally, an output video containing all the essential information about road traffic analysis will be generated. The system will be able to provide real-time traffic data, allowing authorities and drivers to make informed decisions on routes, reducing congestion. With the availability of real-time data, authorities can predict future traffic patterns and manage resources effectively, such as deploying traffic police or adjusting traffic signals.The system will be cheaper and more sustainable in the long run compared to manual methods.

This paper proceeds as follows. In the next Section II, we present the related works, the object detection and object tracking algorithms. The methodology is described in Section III. The implementation, testing, results and evaluation are presented in Section IV. Section V concludes the paper.

II. Literature Review

This section presents a comprehensive overview of the existing research on vehicle counting, road traffic density evaluation and vehicle speed estimation. Mandal and Adu-Gyamfi [6] built algorithms to count and track vehicles. They devel-

oped a reliable approach for vehicle counting on highways and addresses the occlusion issue of small vehicles behind trucks. CenterNet, Detectron2, YOLOv4 and EfficientDet were the four object detectors that were used. Intersection over union (IOU) tracker, Simple Online and Real-time Tracking (SORT), Feature Based Object Tracker were among the object trackers that were tested but Deep SORT was found to be the most reliable one. The combination of CenterNet and Deep SORT outperformed the rest by achieving an average counting accuracy of 95%. Zhu et al. [7] developed an enhanced Single-Shot Detector (SSD) capable of outperforming popular object detectors. ResNet achieved the best results with a classification accuracy of 96.11%, while the accuracy for Speeded-up Robust Features (SURF) was 45.74%. Consequently, a ResNet with reduced layers was picked to be the base network of the improved SSD. An overall accuracy of 90.3% was obtained for the counting and classification of vehicles. This was 3% higher than the one obtained by the traditional SSD.

Song et al. [8] developed a system to conduct vehicle detection and counting on different highways. Firstly, video frames go through the meanshift algorithm and Gaussian filters for smoothing and enhancement. Then, the results are fed to a flooding filling algorithm to identify and extract only the road surfaces. The YOLOv3 algorithm was used to locate the positions of vehicles and the Oriented FAST and Rotated BRIEF (ORB) algorithm was employed to predict driving directions. The experiments revealed that the accuracy for the counting of vehicles and driving directions were 93.% and 92.3%, respectively. Lira et al. [9] constructed a system where video footages obtained from drones were used to perform vehicle detection by using the Mixture of Gaussian (MOG2) algorithm. However, tracking of motorbikes and distant vehicles proved to be difficult. An overall accuracy of 64% was achieved for vehicle detection. Satyanarayana et al. [10] proposed a distinct approach to detect and classify vehicles from Indian roads. The CNN used was trained with 8000 images. The shapes formed by the ROIs (Regions of Interest) were used to approximate the length, width and type of vehicle. However, occurrences of closely moving two-wheelers being classified as a single vehicle of another class were observed. Also, vehicles having colours similar to roads created some difficulties for the system.

Hasan et al. [11] developed a model based on a Convolutional Neural Network (CNN) to analyze traffic density. The CNN was trained to identify five different classes ranging from "empty lanes" to "traffic not moving". One of the worrying issues of the system was overfitting. Data augmentation and batch normalization were used to reduce overfitting. The model was found to be working at an accuracy of 84.06%, with a further increase of 2.5% after batch normalization. Biswas et al. [12] constructed two automated Python algorithms for counting vehicles and for estimating traffic density. The first method was based on Single Shot Detector (SSD) while the other one made use of MobileNet SSD. SSD was able to achieve a detection accuracy of 92.9% while MobileNet only reached 79.3%. Huy and Duc [13] came up with a real-time traffic density evaluation system which locates, counts and classify vehicles using faster R-CNN (Region-Based Convolutional Neural Network) and CSRT (Channel and Spatial Reliability Tracker) algorithm. The tracker can tolerate certain unforeseeable movements and can work with intermittent frame drops. To estimate traffic density, factors such as speed of moving traffic and number of vehicles present were considered. An accuracy of about 95% on highway streets and a 5% drop on crossroads due to frequent occurrences of unpredictable movements was obtained. Bidwe et al. [14] worked on a convolutional neural network (CNN) to classify traffic images into three traffic density categories, namely, low, medium and high. The CNN was able to achieve classification accuracy of up to 99.6%. However, the system can only find density of a single image and has not been tested to calculate the overall density of a sequence of frames.

Ijeri et al. [15] came up with a traffic control system using image processing. Vehicles came from four different directions, the system determines the traffic density and picks the lane that gets green signal for an allocated time. Canny edge detection algorithm is part of the method employed to estimate traffic density. The system is found to be demanding for real-time traffic control. The overall accuracy of the system was only 45%. Mittal et al. [16] put together a traffic density estimation model intended to be used for improving green lights timings at intersections. A combination of Faster R-CNN and YOLOv5, named as EnsembleNet, was used as the object detector. The model was able to operate in low light scenarios and achieve a detection accuracy of up to 98% compared to 95.8% for YOLOv5 and 97.5% for Faster R-CNN for the same test set. Hasanah et al. [17] implemented a Smart Traffic light management system at road intersections in Indonesia. The Haar Cascade Classifier and an advanced CNN were used together to form an object detector. The algorithm can process footages of one second from CCTV cameras in 0.52s for light traffic and 1.05s for heavy traffic. The detection accuracy was 82% during calm periods but drops to 60% during rush hours.

Fedorov et al. [18] attempted to determine the flow of traffic at the most hectic intersection in the Russian city of Chelyabinsk. The faster R-CNN detector was used for vehicle identification and the SORT (Simple Online and Real Time Tracker) algorithm would associate objects throughout the video. Testing revealed an error rate of 7.25% as the tracker struggles when cars overlapped while waiting at the centre of the intersection. Grents et al. [19] also used the architectural model of the faster R-CNN and SORT tracker to detect and track moving vehicles. The authors also estimated vehicles' speed to understand traffic flow using equations from Makwana and Goel [20]. However, the proposed method could only estimate speed within an error rate of 22%. Using YOLOv4, Khalaf et al. [21] achieved an impressive accuracy of 98.93% for the detection of pedestrians from the KITTI dataset

Despite all the breakthroughs concerning vehicle counting and traffic density estimation, a single setup that can compute all the key aspects of road traffic analysis is still missing. Our plan is to implement a system that is able to detect, track, count moving vehicles, estimate traffic flow, traffic density and driving speed. The proposed model should be able to obtain excellent classification accuracy while operating in real-time.

A. Object Detection Algorithms

In the realm of real-time traffic detection systems, object detection stands as a pivotal component, given its fundamental role in identifying every moving vehicle traversing the roadways. While humans effortlessly perform object detection,

machines encounter significant complexity in executing this task. This section delves into recent advancements in object detection methodologies, shedding light on notable object detectors.

1) The YOLO algorithm: At the University of Washington, Joseph Redmon and Ali Farhadi built the start of the well-known YOLO (You Only Look Once) object recognition and picture segmentation model [22]. After its initial introduction in 2015, YOLO has grown in popularity due to its fast speed and correctness [23]. The first model receives frames at a resolution of 448 × 448 pixels and applies a convolutional neural network on the whole image in one go. The network divides the input in a K × K grid, detects objects when their centres are found inside a cell and predicts the positions of bounding boxes and the likelihood of their respective classes. A threshold value is used to reject bounding boxes with low confidence score. The primary YOLO model recorded real-time processing speed of 45 frames per second and a faster version achieved a speed of 155 frames per second [22].

A year later, Redmon and Farhadi released an improved version known as YOLOv2 which implemented a technique to perform training on both object classification and identification [24]. Batch normalization was another addition to improve the precision. YOLOv2 was able to operate at 67 fps and was more robust to different sizes of input images [24]. YOLOv3 consisted of subtle upgrades which made the network layers larger and resulted into better accuracy while not impacting the speed significantly [25]. The following release, YOLOv4, brought the CSPDarknet53 as its classifier backbone along with a 10% increase in speed and launched an innovative data augmentation technique [26]. YOLO's capacity was further enhanced in 2021 with the introduction of YOLOv5 by Ultralytics, which incorporated support for panoptic segmentation and object tracking.

YOLOv8 was published by Ultralytics in 2023 [23]. This model includes a new backbone network as well as a new loss function. One of its attributes is compatibility with all the existing YOLO versions. Users may swiftly switch between and assess the performance of various versions [23]. YOLOv8 can now also perform instance segmentation and image classification in addition to object recognition. Anchor free detection is an architectural change brought into the algorithm to reduce the unnecessary large amount of predicted bounding boxes. Terminating mosaic data augmentation after a certain number of epochs is another update to the system. During testing on the COCO dataset, YOLOv8 had a 20% increase in mAP for its small version compared to YOLOv5 [27].

2) Single Shot Multibox Detector (SSD): SSD (Single Shot Multibox Detector) is a state-of-the-art object detection model. Its architecture consists of only one deep neural network. During training, SSD requires input frames and labelled boxes surrounding each object. When analyzing images, the algorithm's convolutional layers produce several default boxes at the location of each object on feature maps of different proportions. The system then compares both input's boxes and generated ones to look for matches. Those which are compatible are considered positives and the rest are considered as negatives. This technique has proven to enhance SSD accuracy for low resolution images. During detection mode, the convolutional layers generate a feature map and a 3 × 3 convolution kernel is applied on the map to predict a set of bounding boxes where objects are detected along with the likelihood of their respective classes [28].

When tested on the VOC2007 dataset, SSD is able to attain a mean average precision (mAP) of 74.3% and runs at 59 fps while using 300 × 300 pixels images. A slight increase of 2.6% in mAP was reached for 512 × 512 input frames. Nevertheless, as SSD ignores the data beyond the proposed boxes, it is commonly acknowledged that it is less reliable in recognizing smaller objects [29]. Fig. 1 shows the SSD model.

3) Mask R-CNN algorithm: Mask R-CNN is a framework introduced in 2017 by Facebook AI researchers that extended the Faster R-CNN model [30]. The latter operates in two stages. The initial phase, known as a Region Proposal Network (RPN), places bounding boxes on potential objects according to its concurrent prediction of objects limits and objectness values at each point. The following step is derived from Fast R-CNN, uses RoIPool to retrieve traits from each bounding box and classify them into their respective classes. Bounding-box regression is also done to refine localization and the size of boxes. The features identified from these two steps can be communicated between them for quicker inference. Mask R-CNN adds to the second stage by also performing a high precision segmentation of each detected object in the images. When compared to existing models, Mask R-CNN exhibits a greater average precision and does not significantly increase the overhead of faster R-CNN, but it only runs at 5fps, which makes it unsuitable for real-time applications [30].

4) EfficientDet: Tan et al. [31] noticed that system optimization has grown in significance over the last few years for the field of computer vision. They did in-depth research of neural network architectures in an effort to identify the ideal combination to build an efficient object detector. After evaluation, EfficientNet was chosen as the network backbone. Innovative features such as BiFPN (bi-directional feature pyramid network) and a new compound scaling method were introduced. BiFPN acts as the feature network and makes multiscale feature fusion simple and quick. The compound scaling solution simultaneously adjust the pixel density, depth and breadth of network for better optimization. With only 77 million parameters and 410 billion FLOPs, EfficientDet-D7 is able to obtain an average precision of 55.1% on the COCO dataset. Fig. 2 shows the architecture of EfficientDet.

B. Object Tracking Algorithms

Object tracking algorithms constitute an essential element within intelligent traffic monitoring systems. Their primary objective is to establish correspondences between identical vehicles observed across distinct frames of a recording. A proficient algorithm achieves this objective with minimal computational expenditure while ensuring a consistent and reliable tracking of a vehicle throughout its presence in the monitored scene.

1) DeepSORT: DeepSORT is an extension of the computer vision tracking model SORT (Simple Online Realtime Tracking). The SORT algorithm is a method for numerous objects tracking that emphasizes on efficient and straightforward methods to achieve real-time processing. It employs a blend of well-known methods such as the Kalman Filter and

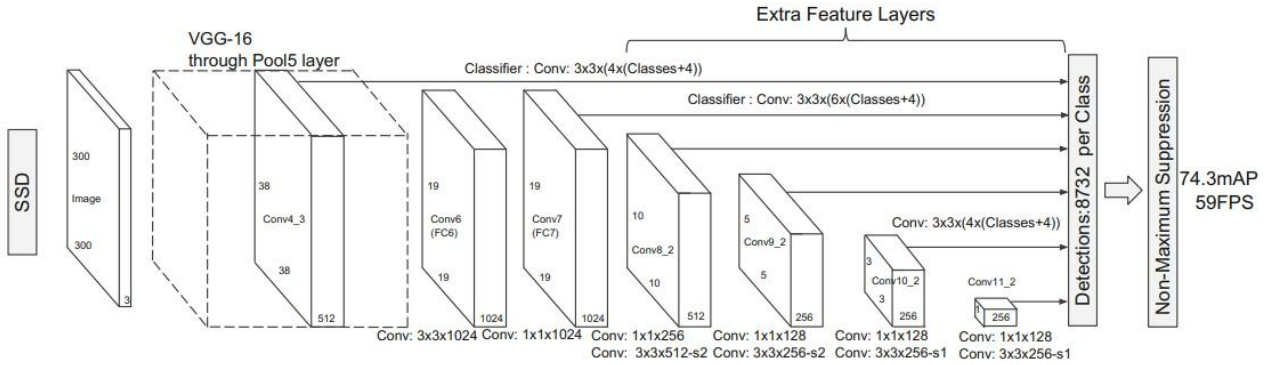


Fig. 1. SSD Model [28].

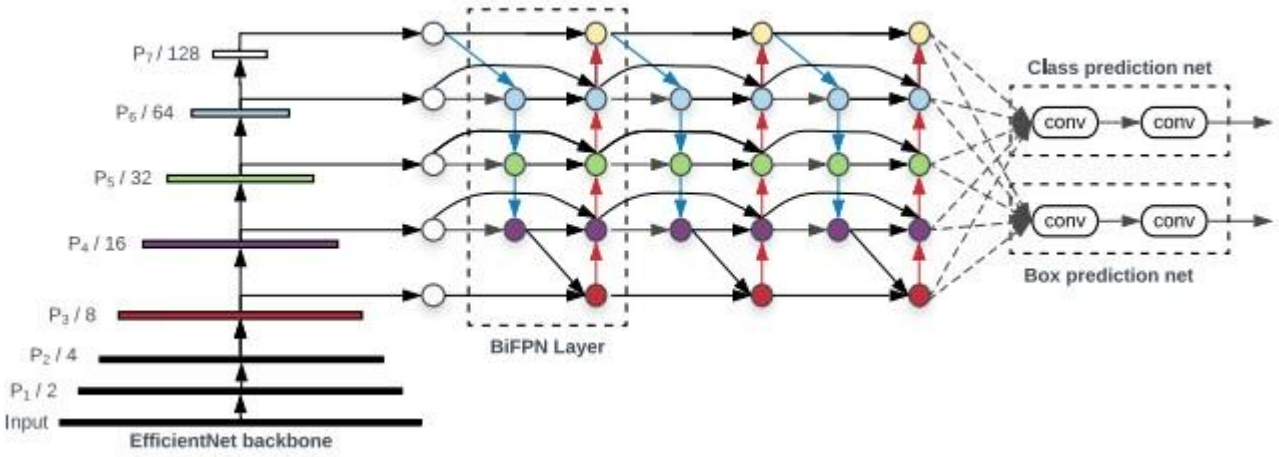


Fig. 2. EfficientDet architecture [31].

Hungarian algorithm to accomplish its goals. When an object appears or exits the frame, a distinct identity is either given to or taken from it. An identified object is labelled as untracked when its detection has an overlap smaller than the defined IOUmin (Intersection over Union) value. An IOU matrix which represents the intersection-over-union distance between each detected object and all expected bounding boxes of pre-existent entities is kept. The matrix is used by the Hungarian algorithm to associate current objects with their next position in the following frame. The Kalman Filter is used to predict future locations of every tracked target which is used for object association. Tracking is terminated if the entity is not found for a specified number of frames [32].

While using the SORT method, some objects' tracking are often lost midway through and afterwards retraced using a new identification. If counting is necessary, it could pose a significant hurdle. DeepSORT tries to include visual appearance of objects as a factor when looking for objects in the upcoming frames. A trained convolutional neural network with two convolutional layers and six residual blocks is incorporated with predicted motion information to form a better object association metric. The technique has shown to eliminate entity switches issue by around 45%. Tracking is proved to be smoother in scenarios with occlusions and unexpected motions [33].

2) ByteTrack: ByteTrack (BYTE) is a simple and robust tracker which main goal is to consider all detected bounding boxes in order to minimize tracking errors due to occluded targets, motion blur or shift in sizes. ByteTrack does not use any convolutional neural network to achieve its aim but rather uses both high and low confidence detection boxes while performing association. Boxes are categorized as low and high confidence according to a selected threshold. The Kalman filter is then applied to predict the expected path of objects in the following frame. The initial association is done with the large confidence value boxes and all the predicted tracks. At this phase, the Hungarian method is utilized to resolve the matching. During the next stage, the resulting unmatched tracks are associated with the low confidence boxes. Any unpaired path is saved so that it may be handled with the following frame. BYTE is a highly customizable model and can easily be merged with other similar tracking algorithms. It was able to obtain an accuracy of 80.3% for multi object tracking while performing at 30 fps [34].

III. Methodology

This section details the architecture and structure of the proposed system. The contents of various elements and how they interact with one another to allow the model to run successfully are listed and thoroughly explained. Before building the components that would be used to extract road traffic information from video recordings, compilation of a robust dataset is essential for training of an object detector.

A. Dataset

The dataset consists of five different classes namely buses, cars, motorbikes, trucks and vans with class ids 0 to 4, respectively. Out of 2800 frames, 22,293 cars, 4852 trucks, 3527 motorbikes, 2308 vans and 1937 buses were manually annotated to form the dataset. Data augmentation techniques such horizontal flips and 10-degree rotations have been used to take the number of frames for the training dataset to a total of 5600 frames. 280 labelled frames were set aside to be used for the validation stage. The images were captured during the day with light, medium and high traffic. Once the videos were captured, any visible faces plate numbers were blurred out as agreed with the Data Protection Office of Mauritius

B. Training of the Model

After intense research and analysis, YOLOv8m has been chosen as the detector to perform vehicle recognition. This choice was made due to the balance of speed and precision it is able to offer. The detector was trained on the dataset for 120 epochs. The process took four hours to be completed.

C. Architectural Framework and Structural Composition of the System

The architecture composes of a detection and classifier module, a tracking unit, a traffic flow computation component, a speed estimation module and a traffic density estimation unit. The detection and classifier module mainly consist of the trained YOLOv8m detector and each video frame, upon being read, is initially transmitted to this detector for processing. It firstly resizes the input image and pass it to its convolutional neural network. Positions of bounding boxes with their corresponding vehicle classes and confidence scores are then determined. Finally, a thresholding of 0.25 is applied to discard low confidence boxes. The output is then transformed into a format compatible with the ByteTrack tracking algorithm and sent to the tracking unit. A reduced ByteTrack algorithm without its association metric serves as the central component within the tracking module. Upon receipt, the detection results are processed by ByteTrack and with the use of the Kalman filter, a list of predicted tracks are created. This list will then be used together with the detection results of the upcoming image for the association of the same vehicle across the sequential video frames. Each track is assigned a tracker ID and paired with the detection exhibiting the highest intersection over union (IOU) score. In case a detection is paired with multiple tracks, only the combination with the highest IOU is considered.

These combinations are then transmitted to the other modules for traffic analysis. The traffic flow computation component is where counting of vehicles is performed. The module requires a reference line perpendicular to the driving direction with end points being pixel coordinates of the recording in order to count vehicles. The end points of the line must be placed in positions such that the visibility of moving targets is optimal while traversing the line. As a vehicle moves from one segment of the line to the other, the counter for its category is incremented. At the end of each 60 s intervals, the number vehicles counted for that period is divided by 60 to calculate traffic flow in vehicles per minute.

The speed estimation module needs two reference lines perpendicular to the driving direction to operate. The calculation of speed for individual vehicles involves determining the time taken to traverse the distance from passing the initial reference line and crossing the subsequent line. The same concept used for vehicle counting is applied to understand when a target is traversing a reference line. Distance between lines can be estimated by adding the length of individual road markings. The traffic density estimation unit makes use of a rectangular box such that two of its edges are perpendicular to the driving direction. The region of interest formed must be located where the detection of vehicle is reliable.

For each 60 s, an average count of vehicles within the specified area in each frame is calculated. This count is divided by the number of lanes and the distance between the edges perpendicular to the driving direction, and then multiplied by 10 to obtain the traffic density per lane in vehicle per 10 m. Fig. 3 shows the activities of the speed estimation module, starting by initializing the speed list to null and positioning the two reference lines needed to calculate the speed of vehicles. Through an examination of tracker IDs associated with recognized vehicles, a distinction is made between new and previously identified vehicles. Newly identified vehicles have their coordinates stored for future reference, while those previously recognized undergo a comparison between their current positions and their last known positions. This assessment aims to ascertain whether these vehicles have crossed either the initial or subsequent reference lines. When a tracker ID is observed to have traversed both lines sequentially, the average speed for that specific duration is computed and then recorded in the speed list.

Fig. 4 describes the workings of the traffic density estimation unit, beginning by defining the coordinates of the rectangular area within which the road traffic density will be estimated. Within each frame, the centroids of all identified vehicles are computed, and their presence within the designated region of interest is determined. Consequently, the traffic density for that particular frame is evaluated based on the count of vehicles within this region. Over a span of 60 s, an average traffic density per lane is calculated.

D. ByteTrack

The instance takes a list of important arguments that have huge impact on its precision and computation speed. track_thresh which is set to 0.25, defines the minimum threshold needed so that an object and a track can be considered a match. track_buffer represent the number of frames a track is kept alive without being updated. The assignment of this parameter, chosen to be 30, necessitates meticulous consideration, particularly in situations characterized by recurrent occlusions. match_thresh is the value used to know if two

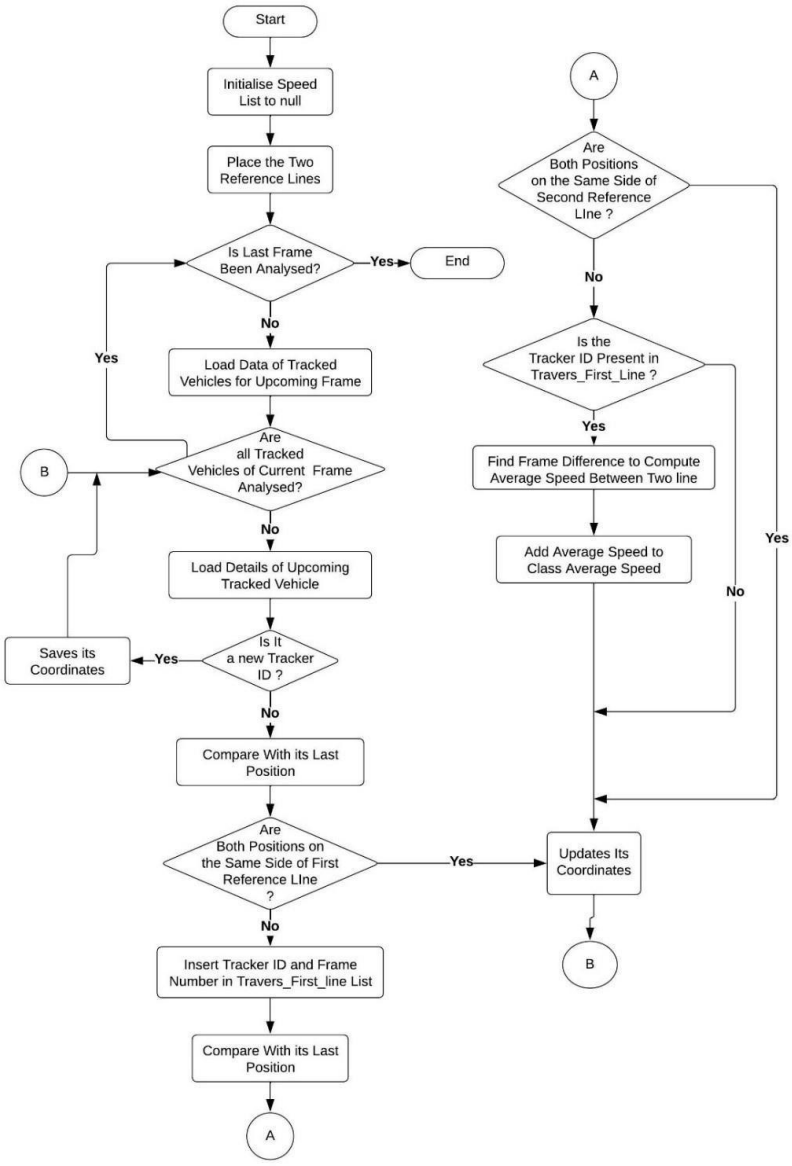


Fig. 3. Speed estimation module.

detections on a frame are the same object and it is configured to a value of 0.8. aspect_ratio_thresh define maximum aspect ratio difference two objects can have to be considered same and it is set to be 3.0. min_box_area, chosen to be 1.0, refers to the minimum area bounding boxes should have to be considered for tracking.

Each detected and tracked vehicle is defined by the co-ordinates of the four corners of its bounding box together with a tracker ID. Whenever a tracker ID is identified for the first time, the x-y coordinates of the corners are tested in a mathematical calculation to find out if they are below or above the line. The Boolean responses are stored in a dictionary. Upon coming across the tracker ID in an upcoming frame, the mathematical calculation is performed again to look for any changes. If all Boolean responses are identical, we wait until the tracker ID may be encountered again. In a situation where only two responses have changed, it would imply that only part of the vehicle has crossed the line. If all four Boolean values have changed, this indicates that the vehicle has crossed the line and is on the opposite side. Consequently, the new values are stored in the dictionary so that the vehicle is not re-identified as crossing the line whenever the tracker ID pops up again.

For each tracked bounding boxes in each frame, its centre point is found by averaging the positions of its four corners. Then, the method cv2.pointPolygonTest from OpenCV is used with the rectangular region and centre point as parameters to find if the point is inside the area. In case of a positive outcome, the vehicle represented by the bounding box is considered to be inside the rectangular box.

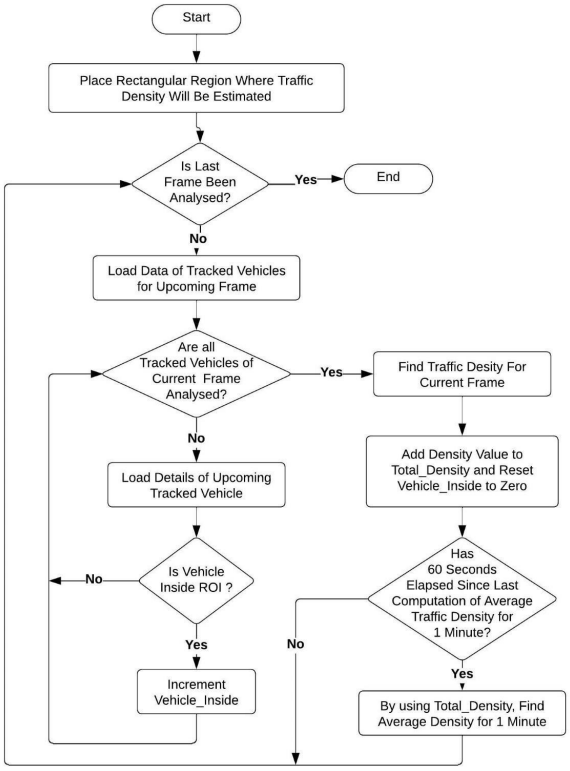


Fig. 4. Traffic density estimation unit.

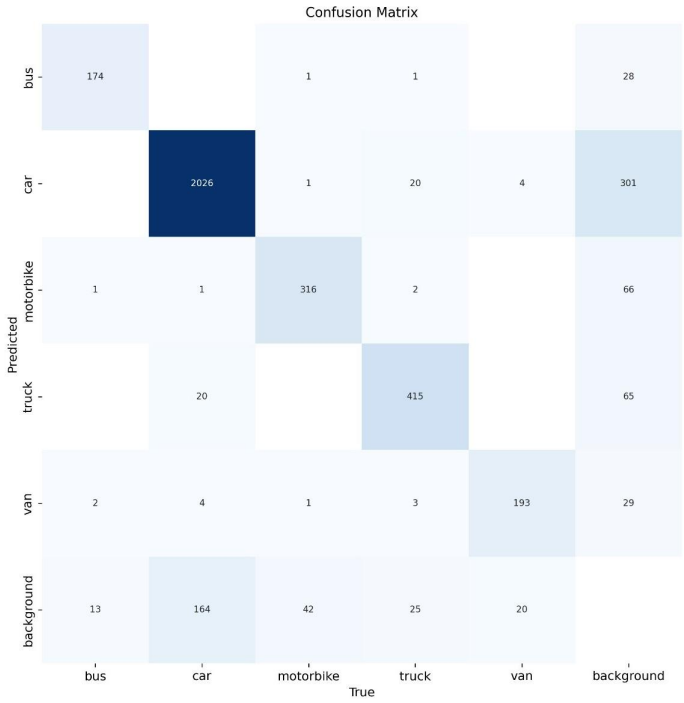


Fig. 5. Confusion matrix.

E. Validation Stage

During the dataset annotation for training purposes, 280 labelled images were set aside to be used for validation. These images were captured under identical conditions and locations as the images used for training. Fig. 5 shows the confusion matrix derived for the validation process. These matrices portray the model's capability to accurately detect and classify 90% of the vehicle instances within the frames. Notably, out of 3429 instances, 486 false positives and 264 false negatives were recorded. The model demonstrates a comparatively higher reliability in identifying buses, while being slightly less accurate for vans.

F. Testing

Test Cases used in this section include scenarios with different road traffic intensities at various locations in Mauritius. The car category includes Support Utility Vehicles (SUV) and the truck category includes lorries, and pickup trucks. The remaining categories are buses, motorbikes and vans.

G. Vehicle Counting and System Speed Comparison

The tracking and counting module have also been built on YOLOv8n and YOLOv8x so that fair accuracy and inference speed comparisons can be made among the three versions. The nano and extra-large versions have also been trained for 120 epochs on the same labelled dataset. From all the videos in test cases, a manual count of vehicles for each category has been done. Classification accuracy has been computed by dividing the count for each vehicle class by their corresponding manual count, find the average and then multiply by 100. The speed of each model has been estimated by the rate at which it can go through a video frame. The reference line for counting is also placed at an optimal position to obtain realistic feedback on correctness.

IV. Experiments, Results and Evaluation

A. Test Case 1: M1 Motorway Near Bagatelle Mall

A two minute video sequence was captured on the bi-directional M1 motorway from the bridge near Bagatelle Mall. The traffic driving to the north had a high traffic density while the traffic moving to the south was sparser. The footage was captured at around 8:30 in the morning. Tables I and II show these results

The results from both directions show that all the three YOLO versions are able to achieve excellent counting and classification accuracies. The traffic flow in both directions for the whole video and for each interval of 60s has also been determined by the system. These values are compared with the actual traffic flow as shown in Tables III and IV.

Observations from Tables III and IV suggest that all the three versions of YOLO are able to estimate traffic flow with very high accuracies. Fig. 6 shows a screenshot from the traffic video used for test case 1.

B. Test Case 2: Brabant Street at Port-Louis

A three minutes video sequence was captured on the bi-directional Brabant Street from the flyover. The road to the north consist of only one lane and had a slow moving traffic while the one to the south consist of two lanes and also had

TABLE I. RESULTS FOR DRIVING DIRECTION: NORTH (TEST CASE 1)

	Manual Count	YOLOv8m	YOLOv8n	YOLOv8x
Car	158	157	158	156
Motorbike	10	10	10	10
Bus	5	5	5	5
Truck	24	22	22	23
Van	10	10	10	10
Total	207	204	205	204
Car Classification Accuracy (%)		99.4	100	98.7
Motorbike Classification Accuracy (%)		100	100	100
Bus Classification Accuracy (%)		100	100	100
Truck Classification Accuracy (%)		91.7	91.7	95.8
Van Classification Accuracy (%)		100	100	100
Counting Accuracy (%)		98.6	99.0	98.6
Classification Accuracy (%)		98.2	98.3	98.9
Speed (FPS)		7	18	3

TABLE II. RESULTS FOR DRIVING DIRECTION: SOUTH (TEST CASE 1)

	Manual Count	YOLOv8m	YOLOv8n	YOLOv8x
Car	46	48	48	45
Motorbike	8	8	8	8
Bus	1	1	1	1
Truck	11	9	10	13
Van	7	7	6	6
Total	73	73	73	73
Car Classification Accuracy (%)		95.7	95.7	97.8
Motorbike Classification Accuracy (%)		100	100	100
Bus Classification Accuracy (%)		100	100	100
Truck Classification Accuracy (%)		81.8	90.9	81.8
Van Classification Accuracy (%)		100	85.7	85.7
Counting Accuracy (%)		100	100	100
Classification Accuracy (%)		95.5	94.5	93.1
Speed (FPS)		7	18	3

TABLE III. TRAFFIC FLOW RESULTS FOR DRIVING DIRECTION: NORTH (TEST CASE 1)

Traffic Flow (Vehicle/min)	Manual Count	YOLOv8m	YOLOv8n	YOLOv8x
0 s–60 s	71	70	71	71
60 s–120 s	68	65	65	65
60 s Intervals Accuracy (%)		97.1	97.8	97.8

TABLE IV. TRAFFIC FLOW RESULTS FOR DRIVING DIRECTION: SOUTH (TEST CASE 1)

Traffic Flow (Vehicle/min)	Manual Count	YOLOv8m	YOLOv8n	YOLOv8x
0 s–60 s	25	24	24	24
60 s–120 s	35	35	35	35
60 s Intervals Accuracy (%)		98.0	98.0	98.0

vehicles moving slowly. The scenario is trickier compared to test case one as a bus stop is found on the road and buses tend to cause vehicle occlusions when they are stationary. This road is also frequently taken by many motorcycles. The footage was captured at around 11:30 in the morning. Tables V and VI show the results.

Fig. 6. YOLOv8m output video.

TABLE V. RESULTS FOR DRIVING DIRECTION: NORTH (TEST CASE 2).

	Manual Count	YOLOv8m	YOLOv8n	YOLOv8x
Car	14	13	15	14
Motorbike	20	19	19	18
Bus	12	12	9	11
Truck	4	1	1	3
Van	1	1	1	0
Total	51	46	45	46
Car Classification Accuracy (%)		92.9	92.9	100
Motorbike Classification Accuracy (%)		95.0	95.0	90.0
Bus Classification Accuracy (%)		100	75.0	91.7
Truck Classification Accuracy (%)		25.0	25.0	75.0
Van Classification Accuracy (%)		100	100	0
Counting Accuracy (%)		90.2	88.2	90.2
Classification Accuracy (%)		82.6	77.6	71.3
Speed (FPS)		7	18	3

TABLE VI. RESULTS FOR DRIVING DIRECTION: SOUTH (TEST CASE 2)

	Manual Count	YOLOv8m	YOLOv8n	YOLOv8x
Car	10	10	10	10
Motorbike	17	17	17	17
Bus	7	6	9	8
Truck	7	5	1	5
Van	1	1	1	0
Total	42	39	38	40
Car Classification Accuracy (%)		100	100	100
Motorbike Classification Accuracy (%)		100	100	100
Bus Classification Accuracy (%)		85.7	71.4	85.7
Truck Classification Accuracy (%)		71.4	14.3	71.4
Van Classification Accuracy (%)		100	100	0
Counting Accuracy (%)		92.9	90.5	95.2
Classification Accuracy (%)		91.4	77.1	71.4
Speed (FPS)		7	18	3

As test case 2 is a more challenging environment, more variations in the results are seen. YOLOv8m outperforms the

two other YOLO versions in classifying the vehicles in their respective categories. Categorization of trucks seems to be more difficult for all the three models. The results from Tables VII and VIII show that all the three versions of YOLO are able to estimate the traffic with accuracies above 80%. Table VII shows a screenshot from the traffic video used for test case 2. Fig. 7 shows the YOLOv8m output video.

TABLE VII. TRAFFIC FLOW RESULTS FOR DRIVING DIRECTION: NORTH (TEST CASE 2)

Traffic Flow (Vehicle/min)	Manual Count	YOLOv8m	YOLOv8n	YOLOv8x
0 s–60 s	13	13	12	13
60 s–120 s	15	10	11	11
120 s–180 s	14	13	11	12
60 s Intervals Accuracy (%)		86.5	81.4	86.3

TABLE VIII. TRAFFIC FLOW RESULTS FOR DRIVING DIRECTION: SOUTH (TEST CASE 2)

Traffic Flow (Vehicle/min)	Manual Count	YOLOv8m	YOLOv8n	YOLOv8x
0 s–60 s	5	4	5	5
60 s–120 s	10	8	7	9
120 s–180 s	21	21	20	20
60 s Intervals Accuracy (%)		86.7	88.4	95.1

Fig. 7. YOLOv8m output video.

C. Test Case 3: Motorway M1 Near Bagatelle Mall

A 4 min video sequence was captured on the bi-directional M1 motorway from the bridge near Bagatelle Mall at around 6 o'clock in the afternoon. Traffic density was very light during this time. Tables IX and X show that all three versions have performed exactly the same while achieving about 99% classification accuracy for both driving directions. The drop in illumination does not seem to have any effect on the detectors' ability to count and classify vehicles.

Tables XI and XII illustrate that all three models had the same performance for traffic flow estimation for each direction. They show that they are able to obtain traffic flow values with accuracies higher than 96%. Fig. 8 shows a screenshot from the traffic video used for test case 3.

In terms of counting accuracy, all the three YOLO versions perform close to each other with YOLOv8x having the slight edge. As predicted, counting results shows that test case 2 was the most challenging as it often had motorbikes and small

TABLE IX. RESULTS FOR DRIVING DIRECTION: NORTH (TEST CASE 3)

	Manual Count	YOLOv8m	YOLOv8n	YOLOv8x
Car	40	38	38	38
Motorbike	5	5	5	5
Bus	2	2	2	2
Truck	6	6	6	6
Van	4	4	4	4
Total	57	55	55	55
Car Classification Accuracy (%)		95.0	95.0	95.0
Motorbike Classification Accuracy (%)		100	100	100
Bus Classification Accuracy (%)		100	100	100
Truck Classification Accuracy (%)		100	100	100
Van Classification Accuracy (%)		100	100	100
Counting Accuracy (%)		96.5	96.5	96.5
Classification Accuracy (%)		99.0	99.0	99.0
Speed (FPS)		7	18	3

TABLE X. RESULTS FOR DRIVING DIRECTION: SOUTH (TEST CASE 3)

	Manual Count	YOLOv8m	YOLOv8n	YOLOv8x
Car	56	55	55	55
Motorbike	1	1	1	1
Bus	0	0	0	0
Truck	3	3	3	3
Van	6	6	6	6
Total	66	65	65	65
Car Classification Accuracy (%)		98.2	98.2	98.2
Motorbike Classification Accuracy (%)		100	100	100
Bus Classification Accuracy (%)		-	-	-
Truck Classification Accuracy (%)		100	100	100
Van Classification Accuracy (%)		100	100	100
Counting Accuracy (%)		98.5	98.5	98.5
Classification Accuracy (%)		99.6	99.6	99.6
Speed (FPS)		7	18	3

TABLE XI. TRAFFIC FLOW RESULTS FOR DRIVING DIRECTION: NORTH (TEST CASE 3)

Traffic Flow (Vehicle/min)	Manual Count	YOLOv8m	YOLOv8n	YOLOv8x
0 s–60 s	17	17	17	17
60 s–120 s	11	10	10	10
120 s–180 s	12	12	12	12
180 s–240 s	15	16	16	16
60 s Intervals Accuracy (%)		96.1	96.1	96.1

TABLE XII. TRAFFIC FLOW RESULTS FOR DRIVING DIRECTION: SOUTH (TEST CASE 3)

Traffic Flow (Vehicle/min)	Manual Count	YOLOv8m	YOLOv8n	YOLOv8x
0 s–60 s	20	19	19	19
60 s–120 s	18	18	18	18
120 s–180 s	16	16	16	16
180 s–240 s	9	9	9	9
60 s Intervals Accuracy (%)		98.8	98.8	98.8

Fig. 8. YOLOv8m output video.

cars being hidden behind large buses waiting at a bus stop. This poses major difficulties for detection and tracking. As for classification accuracy, clear differences can be easily seen among the different models. YOLOv8m achieved a classification accuracy of 94.4% while the accuracy for YOLOv8x was only 88.9%. Test case 2 proved to be challenging for YOLOv8m and YOLOv8x as they misclassified or missed buses and trucks.

As for the operating speed, YOLOv8n outperforms the two others with an average fps of 18 as it is a much lighter detector with fewer parameters. YOLOv8n is more than twice faster than YOLOv8m and almost four times faster than YOLOv8x. The slight variation in speed between different test cases is due to dependency on the number of objects detected. This testing shows YOLOv8m can operate in real-time if a powerful GPU is used. All the three YOLO models are capable of evaluating traffic flow with good accuracies. YOLOv8x is slightly more suited for this task with an overall accuracy of 95.4%.

D. Traffic Density

With a known distance provided for the length of the region of interest (ROI), the system is able to determine traffic density for each lane. Values for traffic density per lane in the ROIs has been manually estimated by counting the number of vehicles present inside the selected areas, divides this figure by the length of the region and the number of lanes then multiplied by ten.

1) Test Case 4: Motorway M1 near bagatelle mall: A three minutes video was recorded on the bi-directional M1 motorway from the bridge near Bagatelle Mall at 9 o'clock in the morning. The video had about 106 vehicles moving north along with about 55 vehicles going south. Both directions consisted of three lanes. A region with a length of sixty was chosen as the length to be inspected for both directions. Tables XIII and XIV show the results.

Values obtained from test case 4 show that the system is able to estimate the traffic density with high accuracies. Fig. 9 shows the outputs from test case 4.

2) Test Case 5: Brabant street at port-louis: A three minutes video was captured on the bi-directional Brabant Street from the flyover at 11 o'clock in the morning. The road to the north consist of only one lane and while the one to the south consist of two lanes. The video had about 18 vehicles moving north along with about 31 vehicles going south. A region with a length of thirty meters was chosen as the area to be inspected for both directions. Tables XV and XVI show the results.

TABLE XIII. RESULTS OF DRIVING DIRECTION TO THE NORTH (TEST CASE 4)

Interval	Traffic Density per Lane Manually Evaluated (Vehicles/per Lane/10 m)	YOLOv8m (Vehicles/per Lane/10 m)	Accuracy (%)
0 s–60 s	1.80	1.77	98.3
60 s–120 s	1.99	1.94	97.5
120 s–180 s	2.11	2.30	91.0

TABLE XIV. RESULTS OF DRIVING DIRECTION TO THE SOUTH (TEST CASE 4)

Interval	Traffic Density Per Lane Manually Evaluated (Vehicles/per Lane/10 m)	YOLOv8m (Vehicles/per Lane/10 m)	Accuracy (%)
0 s–60 s	1.14	1.17	97.4
60 s–120 s	1.00	1.06	94.0
120 s–180 s	0.90	0.89	98.9

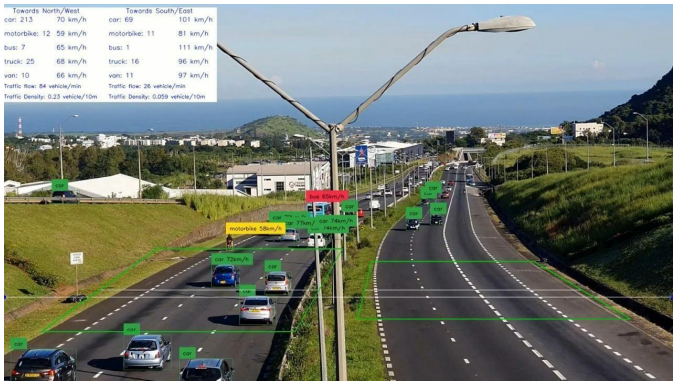


Fig. 9. Output video frame (Test Case 4).

TABLE XV. RESULTS OF DRIVING DIRECTION TO THE NORTH (TEST CASE 5)

Interval	Traffic Density per Lane Manually Evaluated (Vehicles/per Lane/10 m)	YOLOv8m (Vehicles/per Lane/10 m)	Accuracy (%)
0 s–60 s	1.86	2.00	92.5
60 s–120 s	1.61	2.02	74.5
120 s–180 s	2.46	2.55	96.3

For the north direction, during the interval 60 s–120 s, a road marking was wrongly identified as a car for a brief period of time which caused a significant decrease in the accuracy. Fig. 10 shows the outputs from test case 5.

3) Test Case 6: M1 Motorway near bagatelle mall: A three minutes recording was also captured on the bi-directional M1 motorway from the bridge near Bagatelle Mall at around 6 o'clock in the afternoon. The video had around 22 vehicles moving north along with about 30 vehicles going south. A

TABLE XVI. RESULTS OF DRIVING DIRECTION TO THE SOUTH (TEST CASE 5)

Interval	Traffic Density Per Lane Manually Evaluated (Vehicles/per Lane/10 m)	YOLOv8m (Vehicles/per Lane/10 m)	Accuracy (%)
0 s–60 s	1.22	1.26	96.7
60 s–120 s	1.20	1.24	96.7
120 s–180 s	2.71	2.71	100

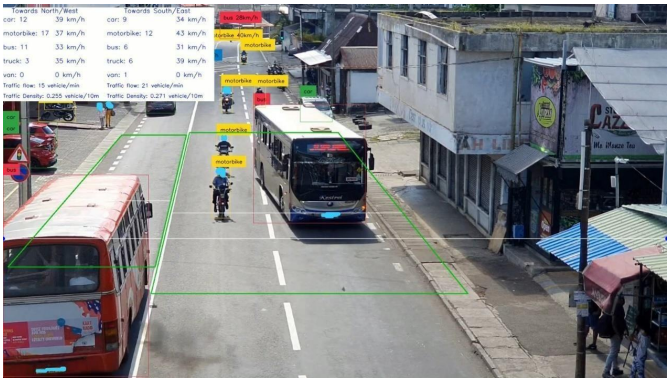


Fig. 10. Output video frame (Test Case 5).

region with a length of forty meters was chosen as the area to be inspected for both directions. There are 3 lanes in both directions. Tables XVII and XVIII show the results. Fig. 11 shows the outputs from test case 6.

TABLE XVII. RESULTS OF DRIVING DIRECTION TO THE NORTH (TEST CASE 6)

Interval	Traffic Density per Lane Manually Evaluated (Vehicle/per Lane/10 m)	YOLOv8m (Vehicle/per Lane/10 m)	Accuracy (%)
0 s–60 s	0.80	0.80	100
60 s–120 s	0.40	0.42	95.0
120 s–180 s	0.62	0.64	96.8

TABLE XVIII. RESULTS OF DRIVING DIRECTION TO THE SOUTH (TEST CASE 6)

Interval	Traffic Density Per Lane Manually Evaluated (Vehicle/per Lane/10 m)	YOLOv8m (Vehicle/per Lane/10 m)	Accuracy (%)
0 s–60 s	0.98	1.00	98.0
60 s–120 s	0.90	0.88	97.8
120 s–180 s	0.64	0.68	93.8

According to the figures obtained from the three test cases, the implementation with YOLOv8m is able to produce an overall mean accuracy of 95.3% for traffic density. An automated real-time traffic monitoring system using computer vision and deep learning offers several advantages. It provides immediate traffic data for better decision-making and congestion management. With automated vehicle counting and tracking, the system ensures higher accuracy and reduces human error. Data obtained from this system can be used by the relevant authorities for decongestion strategies. The system can also be

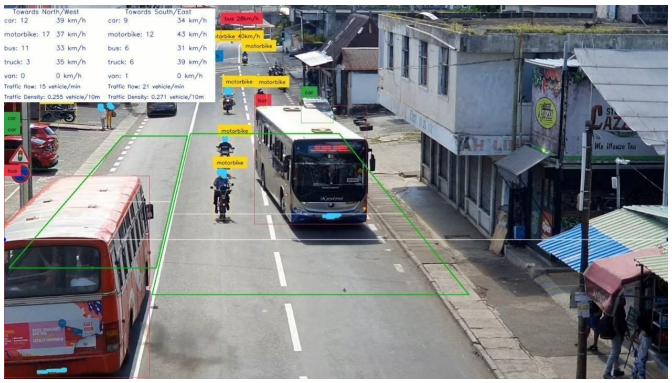


Fig. 11. Output video frame (Test Case 6).

used to enhance road safety by identifying dangerous driving patterns and enable faster emergency responses. Since less time is spent in traffic, it contributes to environmental sustainability by reducing vehicle emissions. For drivers, the system improves their overall driving experience by minimising travel time and stress.

V. CONCLUSION

The recent changes in road infrastructure to incorporate the metro pathway, combined with the constant increase in the number of vehicles in Mauritius, are causing severe traffic congestions on the country's highways and in major towns. The system designed and implemented can undoubtedly serve as the foundation for a solution to this problem. The proposed system is able to detect, track, classify and count vehicles with reasonable accuracy under different traffic conditions. Traffic flow and traffic density can also be estimated in real-time with good accuracy. The system can also estimate the speed of moving targets and give a representation of average driving speed for each category of vehicles. Once deployed, the proposed system can perform these tasks without any human intervention. It is also designed to be scalable and can be integrated with existing monitoring systems to provide detailed information about the traffic state of any roads in the Republic of Mauritius. This will enable the relevant authorities to gain a comprehensive understanding of traffic conditions on any road in Mauritius, facilitating more informed decision-making for road management and future infrastructure projects. As the system evolves, we aim to enhance its robustness by collecting and analyzing additional data from various locations and times, including night-time monitoring, to account for variations in road structures, traffic patterns, and lighting conditions. Future enhancements will focus on refining detection accuracy under challenging conditions, such as poor weather or very heavy traffic, to ensure that the system remains reliable and adaptable for the long-term needs of the transport infrastructure in Mauritius.

REFERENCES

[1] T. Afrin and N. Yodo, "A survey of road traffic congestion measures towards a sustainable and resilient transportation system," *Sustainability*, vol. 12, p. 4660, 2020. [Online]. Available: https://doi.org/10.3390/su12114660

[2] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, 2016, pp. 3464–3468. [Online]. Available: https://doi.org/10.1109/ICIP.2016.7533003

[3] S. Bidwe, G. Kale, and R. Bidwe, "Traffic monitoring system for smart city based on traffic density estimation," *Indian Journal of Computer Science and Engineering*, vol. 13, pp. 1388–1400, 2022. [Online]. Available: https://doi.org/10.21817/indjcse/2022/v13i5/221305006

[4] D. Biswas, H. Su, C. Wang, A. Stevanovic, and W. Wang, "An automatic traffic density estimation using single shot detection (ssd) and mobilenet-ssd," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 110, pp. 176–184, 2019. [Online]. Available: https://doi.org/10.1016/j.pce.2018.12.001

[5] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," 2020, available online: https://arxiv.org/abs/2004.10934v1 (accessed on 23 March 2024).

[6] A. Fedorov, K. Nikolskaia, S. Ivanov, V. Shepelev, and A. Minbaleev, "Traffic flow estimation with data from a video surveillance camera," *Journal of Big Data*, vol. 6, p. 73, 2019. [Online]. Available: https://doi.org/10.1186/s40537-019-0234-z

[7] J. Gorzelany, "The worst traffic jams in history," 2021, available online: https://www.forbes.com/sites/jimgorzelany/2013/05/21/the-worst-traffic-jams-inhistory/?sh=6b5ee2783e1a (accessed on 12 March 2024).

[8] A. Grents, V. Varkentin, and N. Goryaev, "Determining vehicle speed based on video using convolutional neural network," *Transportation Research Procedia*, vol. 50, pp. 192–200, 2020. [Online]. Available: https://doi.org/10.1016/j.trpro.2020.10.024

[9] M. Hasan, S. Das, and M. Akhand, "Estimating traffic density on roads using convolutional neural network with batch normalization," in *Proceedings of the 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, Dhaka, Bangladesh, 2021, pp. 1–6. [Online]. Available: https://doi.org/10.1109/iceeict53905.2021.9667860

[10] M. Hasanah, G. Qorik, O. Pratamasunu, and R. Pawening, "Automatic car detection using haar cascade classifier and convolutional neural network for traffic density estimation," *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)*, vol. 4, pp. 11–18, 2021. [Online]. Available: https://doi.org/10.24014/ijaidm.v4i1.10785

[11] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, pp. 386–397, 2017. [Online]. Available: https://doi.org/10.1109/TPAMI.2018.2844175

[12] T. Huy and B. Duc, "Traffic flow estimation using deep learning," in *Proceedings of the 5th International Conference on Green Technology and Sustainable Development (GTSD)*, Ho Chi Minh City, Vietnam, 2020, pp. 180–184. [Online]. Available: https://doi.org/10.1109/GTSD50082.2020.9303163

[13] D. Ijeri, P. Maidargi, and R. Sunagar, "Traffic control system using image processing," in *Proceedings of the 1st IEEE Bangalore Humanitarian Technology Conference (B-HTC)*, Vijiyapur, India, 2020, pp. 1–6. [Online]. Available: https://doi.org/10.1109/B-HTC50970.2020.9298014

[14] A. Khalaf, M. Abdulrahman, I. Al-Barazanchi, J. Tawfeq, P. JosephMg, and A. Radhi, "Real time pedestrian and objects detection using enhanced yolo integrated with learning complexity-aware cascades," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 22, pp. 362–371, 2024. [Online]. Available: https://doi.org/10.12928/telkomnika.v22i2.24854

[15] G. Li, W. Lai, X. Sui, X. Li, X. Qu, T. Zhang, and Y. Li, "Influence of traffic congestion on driver behavior in post-congestion driving," *Accident Analysis and Prevention*, vol. 141, p. 105508, 2020. [Online]. Available: https://doi.org/10.1016/J.AAP.2020.105508

[16] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. Berg, "SSD: Single shot multibox detector," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Amsterdam, The Netherlands, 2016, pp. 21–37. [Online]. Available: https://doi.org/10.1007/978-3-319-46448-0_2

[17] G. Lira, Z. Kokkinogenis, F. Rossetti RJ, D. Moura, and T. Rúbio, "A computer-vision approach to traffic analysis over intersections," in *Proceedings of the 19th IEEE International Conference on Intelligent Transportation Systems (ITSC)*, Rio de Janeiro, Brazil, 2016, pp. 47–53. [Online]. Available: https://doi.org/10.1109/ITSC.2016.7795530

[18] M. Makwana and P. Goel, "Moving vehicle detection and speed measurement in video sequence," *International Journal of Engineering Research and Technology (IJERT)*, vol. 2, pp. 3534–3537, 2013.

[19] V. Mandal and Y. Adu-Gyamfi, "Object detection and tracking algorithms for vehicle counting: A comparative analysis," *Journal of Big Data Analytics in Transportation*, vol. 2, pp. 251–261, 2020. [Online]. Available: https://doi.org/10.1007/s42421-020-00025-w

[20] U. Mittal, P. Chawla, and R. Tiwari, "Ensemblenet: A hybrid approach for vehicle detection and estimation of traffic density based on faster r-cnn and yolo models," *Neural Computing and Applications*, vol. 35, pp. 4755–4774, 2023. [Online]. Available: https://doi.org/10.1007/S00521-022-07940-9

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779–788. [Online]. Available: https://doi.org/10.1109/CVPR.2016.91

[22] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 6517–6525. [Online]. Available: https://doi.org/10.1109/CVPR.2017.690

[23] ——, "Yolov3: An incremental improvement," 2018, available online: https://arxiv.org/abs/1804.02767v1 (accessed on 3 February 2024).

[24] Road Development Authority, "Budget estimates 2022/23," 2022, available online: https://rda.govmu.org/rda/wp-content/uploads/2022/10/4.-Budget-2022-202312.06.22-1.pdf (accessed on 10 February 2024).

[25] R. Satyanarayana GS, P. Deshmukh, and S. Das, "Vehicle detection and classification with spatio-temporal information obtained from cnn," *Displays*, vol. 75, p. 102294, 2022. [Online]. Available: https://doi.org/10.1016/J.DISPLA.2022.102294

[26] J. Seong, Y. Kim, H. Goh, H. Kim, and A. Stanescu, "Measuring traffic congestion with novel metrics: A case study of six u.s. metropolitan areas," *ISPRS International Journal of Geo-Information*, vol. 12, p. 130, 2023. [Online]. Available: https://doi.org/10.3390/IJGI12030130

[27] J. Solawetz and Francesco, "What is yolov8? the ultimate guide," 2023, available online: https://blog.roboflow.com/whats-new-in-yolov8/ (accessed on 3 February 2024).

[28] H. Song, H. Liang, H. Li, Z. Dai, and X. Yun, "Vision-based vehicle detection and counting system using deep learning in highway scenes," *European Transport Research Review*, vol. 11, p. 51, 2019. [Online]. Available: https://doi.org/10.1186/s12544-019-0390-4

[29] M. Tan, R. Pang, and Q. Le, "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2019, pp. 10 778–10 787. [Online]. Available: https://doi.org/10.1109/CVPR42600.2020.01079

[30] Ultralytics, "Yolov8 docs," 2023, available online: https://docs.ultralytics.com/#ultralytics-yolov8 (accessed on 3 February 2024).

[31] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proceedings of the International Conference on Image Processing (ICIP)*, 2017, pp. 3645–3649. [Online]. Available: https://doi.org/10.1109/ICIP.2017.8296962

[32] W. Xiang, D. Zhang, H. Yu, and V. Athitsos, "Context-aware single-shot detector," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Tahoe, NV, USA, 2018, pp. 1784–1793. [Online]. Available: https://doi.org/10.1109/WACV.2018.00198

[33] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, "Bytetrack: Multi-object tracking by associating every detection box," in *Lecture Notes in Computer Science*. Cham, Switzerland: Springer, 2022, pp. 1–21. [Online]. Available: https://doi.org/10.1007/978-3-031-200472_1

[34] J. Zhu, K. Sun, S. Jia, Q. Li, X. Hou, W. Lin, B. Liu, and G. Qiu, "Urban traffic density estimation based on ultrahigh-resolution uav video and deep neural network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, pp. 4968–4981, 2018. [Online]. Available: https://doi.org/10.1109/JSTARS.2018.2879368

Enhancing Credit Card Fraud Detection Using a Stacking Model Approach and Hyperparameter Optimization

El Bazi Abdelghafour[1], Chrayah Mohamed[2], Aknin Noura[3], Bouzidi Abdelhamid[4]
TIMS LABORATORY, FS Tetouan, Abdelmalek Essaadi University, Tetouan, Morocco[1,3,4]
TIMS LABORATORY, ENSA Tetouan, Abdelmalek Essaadi University, Tetouan, Morocco[2]

***Abstract*—Credit card fraud detection has emerged as a crucial area of study, especially with the rise in online transactions coupled with increased financial losses from fraudulent activities. In this regard, a refined framework for identifying credit card fraud is introduced, utilizing a stacking ensemble model along with hyperparameter optimization. This paper integrates three highly effective algorithms—XGBoost, CatBoost, and LightGBM—into a single strategy to improve predictive performance and address the issue of unbalanced datasets. To enable a more efficient search and adjustment of model parameters, Bayesian Optimization is employed for hyperparameter tuning. The proposed approach has been tested on a publicly accessible dataset. Results indicate notable enhancements over established baseline models in essential performance metrics, including ROC-AUC, precision, and recall. This method, while effective in fraud detection, holds significant promise for other fields focused on identifying rare occurrences.**

***Keywords*—Credit card fraud detection; stacking models; hyperparameter tuning; logistic regression; ensemble learning**

I. Introduction

Identifying fraudulent credit card transactions has emerged as a significant challenge for the financial sector because of the swift growth of digital transactions and online shopping. Current models find it difficult to manage the inherent imbalance in datasets for fraud detection, where fraudulent transactions are infrequent yet expensive. In this regard, every day, finance institutions need to protect billions of transactions happening online. Its fraud prevention horizon, therefore, is at an unimaginable scale. This study seeks to enhance the accuracy of credit card fraud detection by employing a stacking ensemble model along with sophisticated hyperparameter optimization methods. The investigation focuses on these key inquiries: In what ways can stacking models be refined to effectively manage imbalanced datasets? What effect does hyperparameter tuning have on the performance of fraud detection?. Indeed, all estimates point to global financial losses to fraud being in excess of billions of dollars annually, which places a high burden on businesses to create much more improved systems for fraud detection [1]. Security in financial transactions has a critical role in maintaining consumer confidence and the integrity of banking as a whole [2], [3].

The traditional methods involve Logistic Regression, Decision Trees, and Random Forest, among other traditional machine learning techniques that have so far been the bedrock of most fraud detection systems [2]. The way these models work is that they learn from historical data about the patterns which can predict if a transaction is fraudulent or real. Although they have been successful in most cases, the imbalanced nature of fraud datasets, where the ratio of positive instances corresponding to fraudulent activities is much smaller compared to the negative instances corresponding to normal activities, poses a challenge that most traditional models cannot overcome. The class imbalance problem often results in biased models towards the majority class of legitimate transactions, yielding poor detection rates associated with fraudulent activities [4].

Some techniques, therefore, such as SMOTE and undersampling, are often applied to artificially balance a dataset in order to counteract this bias. Although such methods tend to slightly improve the performance of a model, their performance is not always very satisfactory, especially when the relationship in a high-dimensional dataset is complex and non-linear. These challenges have resulted in unprecedented interest in various ensemble learning methods that can achieve better predictive performance by combining multiple classifiers. Ensemble models, more specifically the stacking model, have shown capable performance even better than those of traditional methods at times by exploiting strengths of different algorithms. The mechanism for the operation of stacking models is based on the integration of several base learners, such as Logistic Regression or XGBoost, CatBoost, and LightGBM, with a meta-learner to combine these predictions. In doing so, there can be the realization of a more flexible model that captures different patterns within the data, while improving accuracy and generalization [5], [6]. The efficiency of the stacking model in fraud detection applications follows from the fact that usually, data is high-dimensional, quite complex, with subtle fraud behavior to be captured [7].

Another very important factor for success with machine learning models is hyperparameter tuning. Hyperparameters are generally the most fundamental configurations that regulate the behavior of a machine learning model. Examples include learning rate, depth of decision trees, number of estimators, and many more. Optimizing these hyper-parameter values could highly influence both model accuracy and efficiency. Advanced hyperparameter optimization techniques, especially Bayesian Optimization, Genetic Algorithms, and grid search methods, have over the past couple of years begun to make the process easier by automating and streamlining it. These methods are able to explore the most efficient hyperparameter space with the aim of optimal performance of the model on various datasets [4], [6]. Essentially, it has been found that applying advanced hyperparameter tuning to stacking models

can yield enormous gains in fraud detection performance, typically benchmarked with the ROC AUC score, a popular metric used to measure the effectiveness of the model in distinguishing between fraudulent and legitimate transactions [7].

Fraud Detection Systems require scalability and high-volume data processing in real-time since financial systems generate millions of transactions each day. Since fraudsters continuously evolve their techniques, the machine learning models should be adaptive to keep pace with the detection of emerging trends and patterns in fraud. This means that models always need updates, retraining with new data and tuning generalization performance of the models to unseen scenarios [4]. In this paper the challenges identified are solved by a novel approach considering stacking models combined with an improved hyperparameter optimization inside of a set of base learners: Logistic Regression, XGBoost, CatBoost and LightGBM. This approach will tune the hyperparameters and overcome inherent issues associated with dataset imbalance, thus proving to be more efficient, scalable, and accurate in fraud detection systems by providing a robust solution against fraud transactions within the dynamic digital space.

The organization of the rest of the paper is as follows: Section II discusses related work, considering an in-depth analysis of fraud detection techniques' current status. Section III describes the methodology followed, describing the design of the stacking model along with its optimization process. The results obtained from the experimental evaluation are discussed in Section IV, while Section V concludes the paper and provides insights for future research prospects in this area.

II. Literature Review

A. Overview of Optimization Techniques

The area of optimization has become a stronghold for better performance in machine learning models in different application domains. This often includes tuning the model parameters or hyperparameters to maximize or minimize some objective functions, normally related to model accuracy or error. In large, there are two major levels where optimization techniques are used in machine learning: first, during model training, the adjustment of the parameters to fit the data; second, when doing hyper-parameter tuning, where non-learned parameters like learning rate, number of layers, and so on are tuned for optimal performance of the model [8], [9].

While the field has evolved from traditional approaches like Gradient Descent to other, more sophisticated metaheuristic techniques like Genetic Algorithms (GA) and Particle Swarm Optimization (PSO), developed to deal with various problems such as non-convexity, high-dimensional search space, multi-objective optimization problems [4]. They have been spectacularly improved in domains like fraud detection, healthcare, and financial forecasting, where the data is complex, noisy, and different models are called for based on fine-tuning to achieve high accuracy.

B. Hyperparameter Optimization

HPO is crucial to realize significant improvements in machine learning, which, in application domains such as fraud detection, relies on the accuracy of models. While the neural network parameters are learned during training, the Hyperparameters must be predefined and they limit the learning process, model complexity, and generalization ability [4]. Efficient hyperparameter tuning ensures that the model does not only fit well to training data but generalizes to unseen data, hence improving its overall robustness.

The early models of hyperparameter optimization include Grid Search and Random Search, which systematically search through predefined hyperparameter spaces. Grid Search tests all hyperparameter combinations exhaustively over a specified range, whereas Random Search picks random hyperparameter configurations [9]. While Grid Search ensures that the space is well-covered, this process is computationally expensive, especially for large, high-dimensional datasets, whereas Random Search often outperforms Grid Search in covering a wider area of the hyperparameter space with fewer iterations.

However, both approaches are not very efficient in the case of large hyperparameter spaces, since their convergence to the optimal configuration requires hundreds of evaluations. This consideration motivated the development of more advanced optimization strategies based on intelligent search mechanisms.

Recent development in hyperparameter optimization techniques has derived more sophisticated methods like Bayesian Optimization, Hyperband, and Evolutionary Algorithms [10]. Such methods have been adopted in various domains, including finance and healthcare, since they provide immense power for efficient model optimization in problems involving high stakes, such as fraud detection [11].

C. Overview of Machine Learning

Machine learning, or ML, has been described as one of the most revolutionary technologies in the modern era, deeply embedded in a wide variety of applications that range from segments of healthcare classification [12] and financial domains such as tax collection [13] and cybersecurity [1]. Its ability to process large datasets and find some meaning in them has turned it into an indispensable tool for troubleshooting different problems in reality. The evolution of machine learning models has equipped industries with the power to make data-driven decisions and also unlock predictions at an accuracy level previously unimaginable. ML algorithms are necessarily applied the world over to solve problems in need of precision and scalability, such as fraud detection, risk management, and modeling customer behaviors [7]. Different forms of machine learning have been created to address various kinds of issues.

D. Ensemble Learning

The technique of stacking models, particularly ensemble learning, has become increasingly popular in recent years for enhancing predictive accuracy. This method involves combining various base learners such as Decision Trees, XGBoost, and CatBoost with a meta-learner to achieve improved accuracy and generalization ability [14]. Ensemble learning outperforms different models that work on various applications ranging from fraud detection to credit scoring.

With the development of optimization techniques, computational power, and integration of ensemble learning techniques

over the past few years, the performance and efficiency of machine learning models have improved significantly. This has enabled algorithm stacking or boosting with frameworks such as XGBoost, CatBoost, and LightGBM, allowing models to capture intrinsic, nonlinear relationships within the data that were unprecedented when compared to traditional models, including logistic regression or decision trees [4], [6]. These have also proved to be particularly efficient for high-dimensional, imbalanced datasets that come up in applications related to credit-card fraud detection [5].

E. Logistic Regression

Logistic Regression is among the simplest algorithms of machine learning; it finds its applications, especially in binary classification problems [1]. It calculates the probability of a certain input to belong to one of two classes based on the logistic function. Generally speaking, LR models are linear; that is, there is a linear relationship between input variables and the log-odds of the dependent variable. Due to its simplicity, logistic regression has been widely used in many applications, including medical diagnosis, credit scoring, and fraud detection, since interpretation at low computational cost is possible [4].

However, logistic regression has one major disadvantage: it cannot model data with non-linear trends, which limits its effectiveness on more complex datasets. Hence, logistic regression is common in ensemble models, often used as a meta-learner to combine the predictions of more potent models. It plays a leading role in synthesizing outputs in stacking ensembles, where base learners like decision trees or gradient boosting models create better final predictions. This method has been especially useful in domains like fraud detection, where the relationships among features are highly complex and data are typically imbalanced [1].

F. Gradient Boosting Models

A collection of weak learners in a boosted model collaborates to form a robust predictive model. At each iteration, the model minimizes errors by adjusting for the shortcomings of earlier models. Consequently, such models excel in applications like fraud detection and similar challenges, particularly when dealing with large and imbalanced datasets. Three popular algorithms for gradient boosting are XGBoost, CatBoost, and LightGBM.

1) XGBoost: XGBoost is one of the most efficient and effective techniques that can deal with both classification and regression challenges. It is a scalable machine learning system for faster and accurate predictions [15]. As an example, some of the useful features that make XGBoost very powerful concerns with its capabilities to handle sparse data, missing values since it can handle large datasets with high efficiency, with considerable usefulness in real-world applications such as fraud detection. In XGBoost, sequential trees are created, each of which tries to correct errors of the previously misclassified one. The boosting approach lets it generate high accuracy models and makes them less prone to overfitting, especially when combined with regularizations such as L1 or L2 [4].

XGBoost has gained extraordinary popularity because it is outstandingly competitive in many machine learning competitions, and could be easily scaled to distribute systems and GPUs [15]. Because of its capability to model complicated interaction in imbalanced data, XGBoost was the widely used base learner for fraud detection systems in multilayered models. It is the ensemble technique whose regularization and parallelization features make it particularly fitted to real-life applicative scenarios involving large-scale data in circumstances where swift fraud detection is crucial [5], [14].

2) CatBoost: CatBoost: this is a relatively new-and-awesome gradient boosting algorithm, which gained popularity quite fast, since it handles categorical features better than almost any other algorithm in machine learning. In traditional models, for instance, it was common to perform basic preprocessing either by onehot encoding or label coding so that the representations were numerical values for categorical features; this kind of transformation typically increases the dimensionality of a dataset and introduces some risk related to overfitting . CatBoost handles it through native processing of categorical data that, in turn, reduces preprocessing time and enhances the model.

Probably the most characteristic feature of CatBoost is that it can handle categorical data efficiently, which by default would otherwise need to be converted to one-hot encoding in other machine learning libraries. This makes CatBoost very useful in fields like e-commerce, finance, or fraud detection, where datasets often include naturally categorical variables. Besides this, the CatBoost library uses a lot of special tricks to reduce overfitting and improve the generalization of predictive models, which is highly recommended for large-scale datasets used in the detection of rare events such as fraudulent transactions [7], [14]. Its treatment of imbalanced data further contributed to its impressive performance and hence its popularity in several industries.

3) LightGBM: Probably the most important and challenging problem in machine learning, especially in fraud detection, is that of imbalanced datasets. Usually, fraudulent transactions are a rare class, while more than 95% of transactions are usually legitimate. This leads to biased models toward the majority class. Therefore, essentially, models fail to recognize fraud, and missed fraudulent transactions are often present that result in financial losses. In order to handle it, SMOTE - Synthetic Minority Over-sampling Technique and undersampling techniques are used in order to balance the dataset [1]. Apart from SMOTE, there are scaling techniques, such as Robust Scaler, for standardizing features and reducing the perceived effect of outliers. Robust Scaler scales the features based on interquartile range, which involves less sensitivity to extreme values . This is especially useful for fraud detection datasets, since transaction amounts may have a huge span.

G. Robust Scaler and Imbalance Handling

One of the most challenging tasks for machine learning, especially in fraud detection, is dealing with imbalanced datasets. Fraudulent transactions are generally seldom compared to the normal good ones and often result in models biased toward the majority class. The upshot is that most models thus fail in fraud detection and result in missed fraudulent transactions that translate into financial losses. To address this issue, techniques like SMOTE (Synthetic Minority Over-sampling Technique) and undersampling are used to balance the dataset [1].

III. Methodology

The sections below describe the methodology that was implemented to come up with a workable system for credit card fraud detection. It comprises major steps in methodology such as dataset preprocessing, feature engineering, handling class imbalance, building the stacking ensemble model, and tuning hyperparameters using a Bayesian Optimization Framework. Fig. 1 shows schematic representation of the credit card fraud detection system workflow.

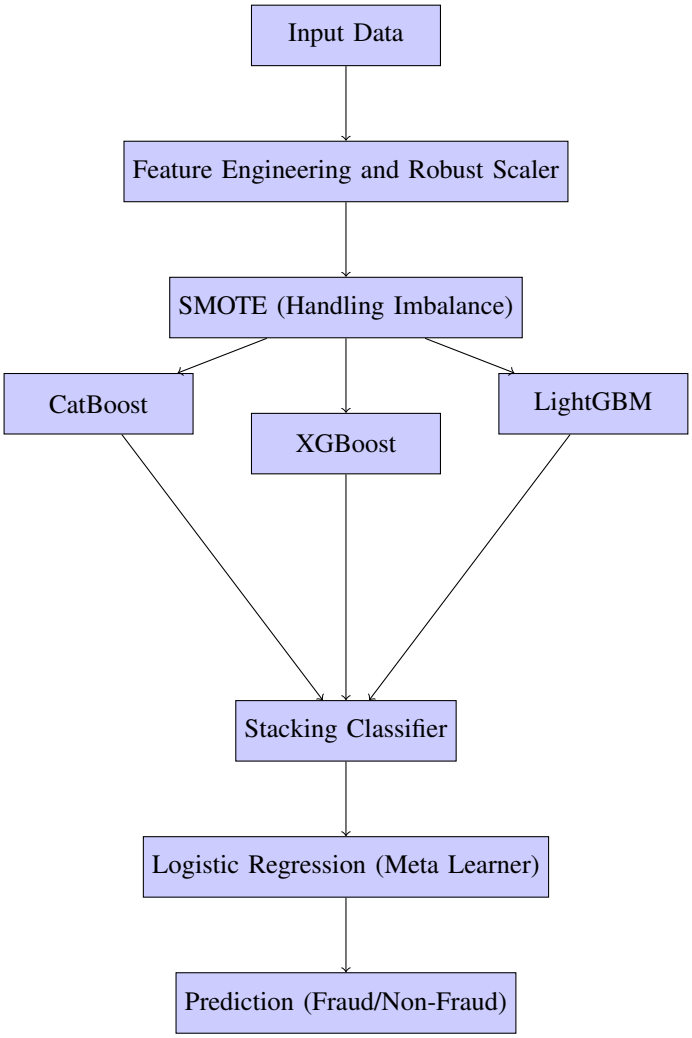


Fig. 1. Schematic representation of the credit card fraud detection system workflow.

A. Dataset Description

Another popular benchmark for fraud detection, the Kaggle Credit Card Fraud Detection dataset [16], is used to benchmark an algorithm due to real-world representation of highly imbalanced datasets of financial transaction data. It contains 284,807 transactions out of which only 492 transactions (approximately **0.172%**) are classified as fraudulent. This highly imbalanced distribution represents real-world scenarios accurately, because the fraudulent transactions occur very rarely, though they bear a high financial impact [1].

In this dataset, there are in total 30 features, from which most of them have been anonymized with Principal Component Analysis - PCA so that individual privacy might not be disclosed. These anonymized features are named V1 to V28, each one corresponding to the principal components obtained from the original data. The other important features of the dataset include the following:

- Time: The time elapsed in seconds since the first transaction in the dataset. This feature can reveal temporal trends, such as specific periods with increased fraudulent activity.
- Amount: The monetary value of the transaction, which can help detect abnormal spending behavior potentially indicative of fraud.
- Class: The target variable, where 1 denotes a fraudulent transaction and 0 denotes a legitimate transaction.

This data distribution is highly imbalanced; most of the transactions here are in the non-fraudulent group. A highly imbalanced, skewed distribution sometimes results in traditional machine learning modeling that is biased toward the majority class, which may not provide the best performance for fraud detection.The distribution of fraudulent vs. non-fraudulent transactions is shown in Fig. 2. Therefore, this can be treated with different techniques such as SMOTE to balance the dataset during training of the model so that it would become more efficient in detecting fraudulent activities.

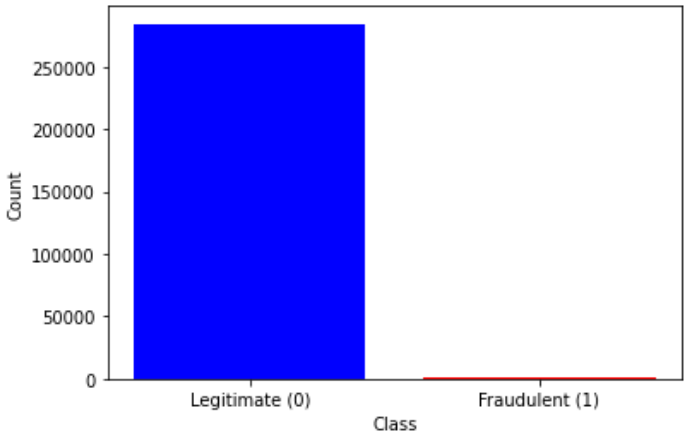


Fig. 2. Distribution of fraudulent vs. Non-fraudulent transactions.

The following is a statistical overview of the Kaggle Credit Card Fraud Detection dataset, summarized in Table I. It focuses on several critical aspects such as the total number of transactions; the number of fraudulent transactions; and the percentage of fraud within the dataset. The high degree of class imbalance may create some problems in getting truly useful models of machine learning.

To enhance the dataset for model training, feature engineering was applied. Key transformations include:

1) Time transformation: The 'Time' feature was converted into an 'Hour' feature, capturing the hour of the day each transaction occurred. This transformation can help reveal temporal patterns in fraudulent behavior.

TABLE I. STATISTICAL INFORMATION OF KAGGLE CREDIT CARD FRAUD DETECTION DATASET

Item	Value
Total Number of Transactions	284,807
Number of Fraudulent Transactions	492
Percentage of Fraudulent Transactions	0.172%
Number of Features (Including Label)	31
PCA Principal Components	28
Transaction Amount Column	1
Time Feature Column	1
Label Column	1

2) Amount transformation: The 'Amount' feature was log-transformed into 'Amount_log' to reduce skewness in the transaction amounts, improving model performance and reducing the impact of extreme values on the learning process [1], [6].

Preprocessing and feature engineering steps are essential to improve the predictive ability of machine learning models , especially when dealing with the uneven nature of the dataset.

B. Handling Class Imbalance

The extreme class imbalance in this dataset has a number of challenges when it comes to the training of machine learning models. If it is not treated properly, the model becomes biased towards the majority class, which consists of all the legitimate transactions, thus making its performance in fraud detection extremely poor. In this work, the Synthetic Minority Oversampling Technique (SMOTE) was used. SMOTE is an algorithm that generates synthetic examples of the minority class through interpolation. This effectively increases the number of fraudulent transactions in the training set without actually oversampling by mere replication of existing records [4]. Moreover, the use of SMOTE is very important for improving the model's performance in correctly classifying fraudulent transactions in high imbalanced datasets [7].

C. Stacking Ensemble Model

To enhance the performance of the fraud detection system, a stacking ensemble model was implemented. Stacking is a method in which several base learner predictions combine using some meta-learner. It helps improve generalization of the model by leveraging each base learner's strengths [4].

The base learners in this stacking ensemble include:

- XGBoost: A highly efficient gradient boosting algorithm known for its speed and ability to handle sparse and large datasets [15].
- CatBoost: Another gradient boosting algorithm that handles categorical features natively without requiring extensive preprocessing, making it particularly suited for datasets with many categorical variables [17].
- LightGBM: Known for its scalability and speed, LightGBM uses a histogram-based approach to efficiently handle large datasets with low memory consumption [18].

These predictions then become the input for the Logistic Regression meta-learner, which then takes this input to give a final classification. Logistic Regression was chosen as the meta-learner due to its simplicity and how effective it is in combining outputs from multiple classifiers in stacking models [4]. The integration of base learners with the meta-learner gives us a robust fraud detection system that can enhance the identification of fraudulent transactions while reducing false positives [5].

D. Hyperparameter Tuning

In order to further enhance the performance of this model, hyperparameter optimization was done using a Bayesian Optimization Framework. It is important to realize that the learning rates, a tree's maximum depth, and the number of estimators are very important hyperparameters, having a great bearing on the effectiveness of any machine learning model. In particular, Optuna is an advanced hyperparameter optimization library that uses Bayesian optimization to dynamically adjust the search space and focus sampling efforts in the most promising areas of hyperparameter space. This allows the approach to be dynamic in its optimization, hence is computationally lighter compared to other approaches based on traditional grid search or random search methods [4], [6].

For each base learner, key hyperparameters were tuned:

- XGBoost: `n_estimators`, `max_depth`, `learning_rate`, `subsample`.
- CatBoost: `iterations`, `depth`, `learning_rate`, `l2_leaf_reg`.
- LightGBM: `n_estimators`, `num_leaves`, `learning_rate`, `feature_fraction`.

Optuna's efficient search process has focused on maximizing the ROC AUC score, a key metric in fraud detection. The ROC AUC score is particularly suitable for evaluating models in imbalanced datasets because it captures the trade-off between true positive and false positive rate across all classification thresholds. Optimized hyperparameters led to stellar improvements while ensuring that the model could cope with the stacking imbalance issue of the dataset with higher accuracy. [4], [5].

E. Model Evaluation and Metrics

Precision, recall, the F1-score, and especially the ROC AUC score were calculated on the final stacking model. In fraud detection, the ROC AUC score is particularly important given that it serves as a measure of balancing between fraudulent and legitimate varies. Better performance is reflected by higher values of AUC. Moreover, metrics such as precision and recall are used to correctly evaluate the performance of the model in terms of reducing false positives while trying to identify fraud. The F1-score offers a balance between precision and recall and is leverage to provide an overall insight into the model's performance on this imbalanced dataset [7].

The ROC curve was plotted to visualize the model's performance across various decision thresholds. The curve provides valuable insights into the trade-offs between sensitivity (recall) and specificity, allowing for an informed selection of the threshold that best meets the system's operational needs.

F. Pseudo Code

In this section, we will outline the pseudo-code for our proposed model.

Algorithm 1 Credit Card Fraud Detection Using Stacking Models and Hyperparameter Tuning

Input: Kaggle Credit Card Fraud Detection dataset
Output: Fraud/Non-Fraud Prediction
procedure MAIN PROCEDURE
    Load dataset *creditcard.csv*
    Feature Engineering:
    Extract 'Hour' from the 'Time' feature
    Apply log transformation on 'Amount' and rename to 'Amount_log'
    Drop original 'Time' and 'Amount' columns
    Data Preprocessing:
    Apply *RobustScaler* on all features except the target ('Class')
    Separate features (X) and target (y)
    Train-Test Split:
    Split the dataset into training and testing sets using *train_test_split*, stratifying by the target 'y'
    Handling Imbalance:
    Use *SMOTE* to oversample the minority class in the training set
    Hyperparameter Optimization:
    Define the *objective* function for optimization:
    Tune parameters for *XGBoost, CatBoost, LightGBM* using *Bayesian Optimization*
    Define parameter ranges: *n_estimators, max_depth, learning_rate*, etc.
    Evaluate model using *ROC AUC score*
    Model Training with Stacking:
    Initialize base models: *XGBoost, CatBoost, LightGBM*
    Define a *Stacking Classifier* with the base models and a *Logistic Regression* final estimator
    Perform 5-fold cross-validation with *StratifiedKFold*
    Model Fitting:
    Train the stacking model using the training data
    Predict fraud probabilities on the test set
    Evaluation:
    Calculate the *ROC AUC score* for model evaluation
    Generate a classification report with precision, recall, F1-score
    Plot and save the *ROC Curve*
    Error Handling:
    If an error occurs, log the error with traceback information
end procedure

IV. Results and Discussion

This section presents the results of our stacking ensemble model on the Kaggle Credit Card Fraud Detection dataset, followed by a detailed comparison with the results from Jiang et al. (2023) [19]. The key performance metrics considered are Precision (PR), Recall (RC), F1-Score (F1), and ROC AUC. Additionally, the ROC curve of our model is provided for visualization of its classification performance.

A. Our Model's Performance

In summary, as shown in Table II, the performance from our stacking ensemble model using Logistic Regression as the meta-learner and base learners of XGBoost, CatBoost, and LightGBM. Our model yields a performance of ROC AUC score of 0.9887 that was strong in the classification and yields good separation between fraudulent and legitimate transactions for many thresholds. This reflects a high score for the model's precision in recognizing both classes of transactions even from an imbalanced dataset where fraudulent transactions are few.

The high Recall score points to the ability of the model to detect a large proportion of fraudulent transactions, something that in fraud detection scenarios is very important because missing fraud cases results in significant financial losses. On the other hand, the model keeps a strong Precision, meaning it finds most of the fraudulent-flagged transactions to be actually fraud. This trade-off between precision and recall is important for minimum false positives, which, in turn, keeps operational costs lower for the review of legitimate transactions that have been wrongly flagged as fraud.

Furthermore, the F1-Score shows the general performance of the model in its balance of the detection of fraudulent and legitimate transactions. This is especially achieved using the stacking ensemble approach, which aggregates the strengths of several machine learning models. In such a method, it allows base learners at different layers to focus on various aspects of the data, improving their capacity for better detection of subtle patterns in fraudulent activities that may get lost with just one model. Meanwhile, it uses the integration or combination of different algorithms: XGBoost, CatBoost, and LightGBM, which then serves as a versatile and adaptive model, since the best performance will be achieved at everything related to the types of transactions.

TABLE II. PERFORMANCE OF OUR STACKING ENSEMBLE MODEL

Metric	Class 0 (Legitimate)	Class 1 (Fraudulent)	Macro Avg	Weighted Avg
Precision	1.00	0.88	0.94	1.00
Recall	1.00	0.86	0.93	1.00
F1-Score	1.00	0.87	0.94	1.00
ROC AUC		0.9887		

Fig. 3 presents the ROC curve further enforces the fact of high discriminatory power of our model. The magnitude of the AUC at a value of 0.9887 was impressive and ensures excellent classification performance over an immense variance in decision thresholds. This high score essentially means that the model is highly effective in discriminating between fraudulent and legitimate transactions by balancing the sensitivity with the specificity of the model. The fact that such high classification performance can be steadily kept with different thresholds is core in guaranteeing that the model can minimize prematurely both false positives and false negatives. For this reason, the system proves to be not only quite accurate but also reliable in real-life applications where fraudulent activities need to be detected with precision while disrupting legitimate transactions as little as possible.

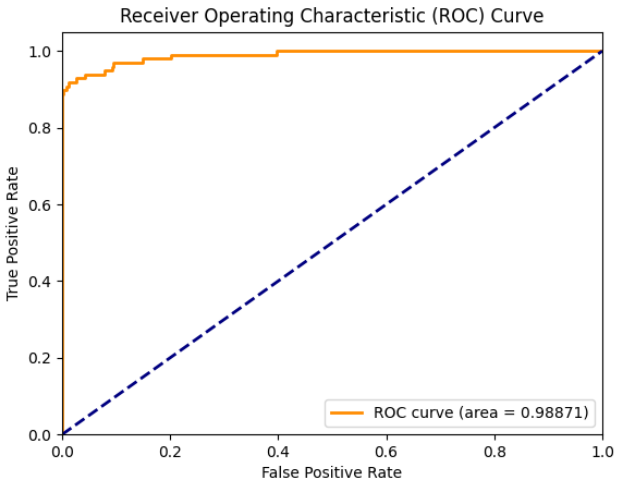


Fig. 3. ROC Curve of our stacking ensemble model

B. Comparison with Jiang et al. (2023)

Apart from the assessment of the performance of our stacking ensemble model, it was relevant to put it in comparison with other state-of-the-art methodologies within this related Research area: credit card fraud detection. Among the rela- To tively new and more sophisticated ones comes the attention Mechanism-based unsupervised anomaly detection network UAAD-FDNet, derived from the work of Jiang et al. (2023) [19]. This model leverages unsupervised learning. combined with attention mechanisms to perform anomaly detection. extend fraud detector performance.

Table III compares the performance of our proposed approach against these existing approaches, including results derived directly from the work of Jiang et al. (2023) [19] for their UAAD-FDNet model and other baseline approaches.

TABLE III. COMPARATIVE RESULTS ON KAGGLE CREDIT CARD FRAUD DETECTION DATASET (RED BOLD INDICATES OPTIMAL RESULTS)

Method	PR	RC	F1	AUC
SVM	0.8854	0.7215	0.7951	0.8586
DT	0.8837	0.7269	0.7977	0.8598
XGBoost	0.8955	0.7280	0.8031	0.8649
KNN	0.9032	0.7268	0.8055	0.8709
RF	0.9112	0.7343	0.8132	0.8827
LSTM	0.9073	0.7391	0.8146	0.8845
CNN	0.9217	0.7453	0.8242	0.9075
MLP	0.9262	0.7461	0.8265	0.9094
AE	0.9528	0.7495	0.8390	0.9279
UAAD-FDNet w/o FA (Jiang et al.)	0.9756	0.7514	0.8489	0.9437
UAAD-FDNet w/ FA (Jiang et al.)	**0.9795**	0.7553	0.8529	0.9515
OptiStack (Ours)	0.88	**0.86**	**0.87**	**0.9887**

C. Discussion of Comparative Results

Table III highlights the comparative performance of our stacking ensemble model and the various models evaluated by Jiang et al. (2023). Our model performs competitively, especially in terms of Recall, F1-Score and ROC AUC, while Jiang et al. (2023)'s UAAD-FDNet model achieves the best performance in terms of Precision.

a) Precision: The UAAD-FDNet w/ FA model in Jiang et al. (2023) was able to achieve an accuracy of 0.9795, while that of our model stood at 0.88. This shorthand form of explanation clearly states that their model returns many fewer false positives and is only suspecting a tiny fraction of legitimate transactions as fraudulent. Precision does retain its value in terms of minimum damage caused to valid users, which the model does excel at.

b) Recall: Our model further outperforms the models of UAAD-FDNet for Recall, having a value of 0.86 versus 0.7553 for UAAD-FDNet w/ FA. This is indicative of the fact that our model can capture a larger portion of fraudulent cases- a critical factor in fraud detection systems, since false negatives (cases of missed fraud) are expensive.

c) F1-Score: is a balance between precision and recall. Jiang et al. (2023)'s UAAD-FDNet w/ FA model has an F1-Score of 0.8529, slightly lower than the 0.87 from our model. That would say, while our model sacrifices some on precision, it balances capture of fraud with low false positives better.

d) ROC AUC: Our model outperformed all models on the ROC AUC score, including Jiang et al.'s UAAD-FDNet models, with the best AUC reached at 0.9515 while our best ROC AUC score was 0.9887. The higher the value of the ROC AUC score, the greater the generalizability of a model for a wide range of decision thresholds when classifying fraudulent and legitimate transactions.

D. Strengths and Areas for Improvement

Our model is very powerful in finding fraudulent transactions, as can be shown by high Recall and ROC AUC. In as far as Precision is concerned, this model has room for further improvement. It is beaten by the UAAD-FDNet model. There is an open challenge in fraud detection where reducing the number of false positives with a high recall is a challenge. The future directions could be the use of hybrid models or further tuning of hyperparameters in a way that precision improves without hurting the recall. Additionally, more sophisticated data augmentation or the usage of fraud detection systems in real time may further improve the robustness of this model. Furthermore, using extra features related to temporal or behavioral patterns may provide the key toward much better overall prediction accuracy. Continuous training of the models with updated fraud patterns can maintain the adaptability of the system against novel fraudulent activities. Last but not least, deeper neural networks or graph-based models could present opportunities for improved results.

V. CONCLUSION

This work has successfully performed a stacking ensemble model that comprises of XGBoost, CatBoost, and LightGBM; deal with Logistic Regression as the meta-model, for credit card fraud detection. Advanced hyperparameter optimization using Bayesian Optimization Framework has been performed very successfully, which yields a very big boost in performance compared to usual single-model solutions.

It shows the ability of the model to discriminate reliably between fraudulent and legitimate transactions, as can be seen by the model's performance, especially the ROC AUC score

of 0.9887. The model's high recall of 0.86 and F1-score of 0.87 demonstrate a critical perspective in fraud detection systems where missed fraud cases should be as low as possible. Although the precision is somewhat lower than that of some competing methods, the overall balance of the model ensures that this can be safely deployed to environments that prioritize fraud identification without overwhelming users with false positives.

Compared to the state-of-the-art UAAD-FDNet from Jiang et al. (2023) [19], our results are competitive and mostly surpassing. Although the approach by Jiang et al. was able to show high precision, ours showed better all-around classification capabilities, especially with recall and AUC, making it highly applicable for fraud detection in real-world settings.

Besides achieving a good performance, the proposed method outperforms others by showing great potential of ensemble models complemented with effective techniques of hyperparameter tuning and handling imbalanced data like SMOTE. These methods address critical fraud detection challenges: class imbalance and precision for detecting rare fraudulent transactions.

Future work may focus on further ehancment of precision, perhaps with even newer and more complicated ensemble methods, or working towards real-time deployment in high-throughput finance systems. Another interesting aproach might be the application of this method on fraud detection problems outside this dataset for verification across different domains.

In the end, out approach of OptiStack issues out to be robust and flexible for detecting credit card fraudulent transactions. The base formed will satisfy advanced stacking methods and hyperparameter tuning, hence potentially helping enhance financial security and combat fraud.

REFERENCES

[1] C. Phua *et al.*, "A comprehensive survey of data mining-based fraud detection research," *Artificial Intelligence Review*, vol. 33, no. 3, pp. 229–246, 2010.

[2] S. Xu, G. Liu, Z. Li, L. Zheng, and S. Wang, "Random forest for credit card fraud detection," *IEEE International Conference on Networking, Sensing, and Control*, pp. 1–6, 2018.

[3] M. Al-shabi, "Credit card fraud detection using autoencoder model in unbalanced datasets," *Journal of Advanced Math and Computer Science*, vol. 33, pp. 1–16, 2019.

[4] E. Esenogho, I. Mienye, T. Swart, K. Aruleba, and G. Obaido, "A neural network ensemble with feature engineering for improved credit card fraud detection," *IEEE Access*, vol. 10, pp. 16 400–16 407, 2022.

[5] J. A. Moses Ashawa, Jude Osamor, "Enhancing credit card fraud detection: An ensemble machine learning approach," *Big Data and Cognitive Computing*, vol. 8, no. 1, 2024.

[6] R. Al-Sulaiman and P. Sant, "Credit card fraud detection using stacking models with enhanced swarm optimization," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no. 3, 2023.

[7] S. Makki, Z. Assaghir, Y. Taher *et al.*, "An experimental study with imbalanced classification approaches for credit card fraud detection," *IEEE Access*, vol. 7, pp. 93 010–93 022, 2019.

[8] J. Nocedal and S. J. Wright, *Numerical optimization*. Springer Science & Business Media, 2006.

[9] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, no. 2, pp. 281–305, 2012.

[10] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," *Advances in neural information processing systems*, vol. 25, 2012.

[11] A. Alshehri and R. Ahmed, "Fraud detection in financial institutions using machine learning algorithms," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 13, no. 11, 2022.

[12] H. Mustafa, C. Mohamed, O. Nabil, and Noura, "Machine learning techniques for diabetes classification: A comparative study," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 9, pp. 785–790, 2023.

[13] N. Ourdani, M. Chrayah, and N. Aknin, "Towards a new approach to maximize tax collection using machine learning algorithms," *IAES International Journal of Artificial Intelligence*, vol. 13, no. 1, pp. 737–746, 2024.

[14] F. Alomari and S. Alqarni, "Credit card fraud detection using stacked machine learning algorithms: A comparative study," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 13, no. 10, 2022.

[15] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

[16] A. D. Pozzolo, "Credit card fraud detection dataset," https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud, 2015.

[17] A. V. Dorogush, V. Ershov, and A. Gulin, "Catboost: gradient boosting with categorical features support," *arXiv preprint arXiv:1810.11363*, 2018.

[18] G. Ke *et al.*, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, vol. 30, pp. 3146–3154, 2017.

[19] M. Jiang, Y. Liu, J. Yin, and C. Xiao, "Credit card fraud detection based on unsupervised attentional anomaly detection network," *AI*, vol. 3, no. 1, pp. 1–15, 2023.

Towards Interpretable Diabetic Retinopathy Detection: Combining Multi-CNN Models with Grad-CAM

Zakaria Said[1]*, Fatima-Ezzahraa Ben-Bouazza[2], Mounir Mekkour[3]
Mathematical Analysis and Applications Laboratory, University Sidi Mohamed Ben Abdellah, Fes, Morocco[1]
Faculty of Science and Technology, Hassan I University, Settat, Morocco[2]
Mohammed VI University of Health Sciences, Casablanca, Morocco[2]
LaMSN, La Maison Des Sciences Numériques, France[2]
Mathematical Analysis and Applications Laboratory, University Sidi Mohamed Ben Abdellah, Fes, Morocco[3]

***Abstract*—Diabetic retinopathy (DR) is a leading cause of vision impairment and blindness, necessitating accurate and early detection to prevent severe outcomes. This paper discusses the utility of ensemble learning methodologies in enhancing the prediction accuracy of Diabetic Retinopathy detection from retinal images and the prospective utilization of Gradient-weighted Class Activation Mapping (Grad-CAM) to maximize model interpretability. Using a dataset of 1,437 color fundus images, we explored the potential of different pre-trained convolutional neural networks (CNNs), including Xception, VGG16, InceptionV3, and DenseNet121. Their respective accuracies on the test set were 89.27%, 91.44%, 89.06%, and 93.35%. Our objective was to improve the accuracy of diabetic retinopathy detection. We explored methods to combine predictions from these four models we began with weighted voting, which achieved an accuracy of 93.95%, and subsequently employed meta-learners, achieving an improved accuracy of 94.63%. These approaches surpassed individual models in distinguishing between non-proliferative and proliferative phases of DR. These findings underscore the potential of these approaches in developing robust diagnostic tools for diabetic retinopathy. Furthermore, techniques like Grad-CAM enhance interpretability, opening the door for further advancements in early-stage detection and clinical integration automatically while maximising accuracy and interpretability.**

***Keywords*—Diabetic retinopathy; retinal images; Grad-CAM; weighted voting; meta-learners**

I. Introduction to Diabetic Retinopathy Detection

A. Background

Diabetes is one of the most prevalent diseases worldwide. It is a chronic condition characterized by elevated blood sugar levels (hyperglycemia) [1]. It involves the assimilation, utilization, and storage disorder of sugars in the diet.It occurs when the body fails to effectively utilize the insulin it produces or when the pancreas does not produce an adequate amount. Insulin is a hormone that is indispensable for the regulation of blood sugar levels by permitting glucose to access the body's cells. Excessive urine excretion, intense thirst, constant appetite, weight loss, impaired vision, and fatigue are among the most prevalent symptoms.In order to prevent long-term complications that affect various body systems, particularly nerves and blood vessels meticulous management is necessary.

According to a report by the World Health Organization (WHO) [2], more than 400 million people are suffering from diabetes in the world. It is anticipated that this figure will rise to 552 million in 2024. The World Health Organization also reports that diabetes is a significant cause of blindness, amputations, and mortality. One of the primary causes of blindness is diabetic retinopathy. More than 5 million people around the world with diabetes are blind. This number is expected to double by 2030. Research at a Jakarta government hospital in 2011 [3] indicated that the highest diabetes complication was neuropathy (54%), followed by diabetic retinopathy (33.4%) in second place.

The relationship between diabetes and diabetic retinopathy is both direct and causal. Diabetic retinopathy is an ocular complication caused directly by diabetes. The chronic hyperglycemia (high blood sugar levels) associated with diabetes damages the microscopic blood vessels in the retina [4]. This can lead to blood leakage, the formation of abnormal new blood vessels, and other alterations in the retina. The risk of developing diabetic retinopathy increases with the duration of diabetes and inadequate glycemic control. The longer the diabetes and the lesser the glycemic control, the higher the risk. Initially, diabetic retinopathy may be asymptomatic. However, as it progresses, it can cause mild to severe vision problems, even leading to blindness. A comprehensive eye examination by an optometrist can enable the early detection of diabetes and diabetic retinopathy, reducing the risk of visual loss. Early detection and effective management of diabetes, including excellent glycemic control, is essential to prevent or delay the progression of diabetic retinopathy. In summary, diabetic retinopathy is a direct and frequent consequence of diabetes, underscoring the significance of its detection and monitoring in preserving ocular health.

Deep learning algorithms have made significant advances in improving diabetic retinopathy detection in recent years, which can help physicians make informed decisions about the most effective treatment plan for each patient. This article will discuss the application of deep learning to detect diabetic retinopathy and the exploitation of some techniques that will help push the models performances in terms of diabetic retinopathy detection and other systems that will improve their results interpretability, It will also cover the performance metrics and validation techniques used to assess the efficacy of

these models. Finally, we will examine how these results may influence future treatment procedures for diabetic retinopathy patients [5], [6].

The following sections of our paper are structured in a way that facilitates clarity and comprehensiveness. Section II provides a comprehensive description of the Dataset we used in our research. In Section III, we outline the approach we took to ensure accurate and reliable results. Section IV discusses the findings of our study and provides a comprehensive analysis of the used methodologies. We then delve deeper into the implications of our findings. Finally, in Section V, we present our conclusions and suggest potential avenues for future research.

B. Literature Review

Research on the classification of Diabetic Retinopathy (DR) has been extensive. Gondal et al. [7] introduced a CNN model with 93.6% sensitivity and 97.6% specificity , using Kaggle and DiaretDB1 datasets. Wang et al. [8] introduced a model that combined different networks and achieved AUC scores of 0.978 and 0.960 . Quelle et al. [9] worked on CNN models for binary classification and lesions detection. Chandrakumar and Kathirvel [10] achieved 94% accuracy on the DRIVE and STARE datasets using a CNN model with dropout regularization. Memon et al. [11] applied nonlocal mean denoising and brightness equalization, achieving a kappa score accuracy of 0.74. Pratt et al. [12] developed a CNN for five DR stages but struggled with mild stage classification due to dataset imbalance. Yang et al. [13] introduced a DCNN for normal and NPDR stages with lesion highlighting and grading. Garcia et al. [14] assessed CNN models (Alexnet, VGGnet16) on the Kaggle dataset, achieving 83.68% accuracy on VGG16. Dutta et al. [15] used the Kaggle dataset to assess three deep learning models, with the best training accuracy of 89.6% on a DNN. Recent advancements include Luo et al. [16], who proposed Multi-View DRD (MVDRNet) combining DCNNs and attention mechanisms, though it failed to train a network with lesion explanation. Chen [17] introduced a multi-scale shallow CNN model for early DR recognition, but it did not significantly improve classification precision. Martinez-Murcia et al. [18] created a CNN for routine DR diagnosis, which was not practical for clinical applications. Deepa et al. [19] created a Deep CNN (MPDCNN) for fundus image recognition, but it lacked advanced neural network architectures for more accurate detection. Das et al. [20] designed a deep learning architecture for DR categorization based on segmented fundus images, while Kalyani et al. [21] introduced a reformed capsule network for feature extraction from fundus images. Oh et al. [22] developed a novel DRD method using top fundus photography and deep learning techniques but did not effectively set an ROI for minimizing complexity. Erciyas and Barısci [23] applied deep learning techniques for automatic lesion detection through ROI extraction, but their method did not optimize system resource utilization.

To summarize, research on DR classification may be categorized into binary and multi-class classification. Binary classification is limited in assessing the severity of Diabetic Retinopathy (DR). On the other hand, multi-class classification categorizes Diabetic Retinopathy (DR) into five distinct phases. Currently used models face challenges with the learning of the abstracted characteristics out of the different stages the diabetic retinopathy and accurately categorizing the stages of diabetic retinopathy in the inference process , which is crucial for successful therapy outcomes.

In response to this, our study aims to identify the stages of diabetic retinopathy accurately. We focused on maximizing the accuracy of the models and providing an interpretability approach to maximize the potential of our approach and assist the medical staff in making decisions based on concrete factors [24], [25], [26].

C. Research Contribution

Our contribution involves a focused effort on enhancing the performance of our models to detect diabetic retinopathy more effectively compared to others addressing the same issue. Our primary goal was to improve the effectiveness of our models in identifying diabetic retinopathy. We adjusted our strategy to enhance the model's ability to differentiate between the three main phases of diabetic retinopathy [27] NDR (No Diabetic Retinopathy), NPDR (Non-Proliferative Diabetic Retinopathy), and PDR (Proliferative Diabetic Retinopathy). Exploring multiple models for the task allowed us to evaluate their effectiveness and refine our approach for greater efficiency.

In the following section, we delve into interpretability and visual assistance concepts. This is relevant because diabetic retinopathy is caused by high blood sugar levels, which degrade capillary walls and result in leakage. This leads to the rupture and bursting of retinal vessels. Our goal is to equip medical professionals with a visualization technique using our advanced deep learning models. This method emphasizes crucial areas to help doctors detect abnormalities and potential lesions by focusing on the regions influencing the model's prediction.

The combination of these two methods enables precise detection and the creation of a comprehensive interpretive framework. This progress will enable the establishment of a system that greatly aids doctors in timely and accurate diagnoses, resulting in improved patient outcomes and a reduced risk of vision loss. Furthermore, this technology promises to streamline the diagnostic process, enabling prompt treatment and intervention as necessary.

II. Dataset and Preparation

To realize our project, we sought a representative database [28], [29] that would encompass the various stages and manifestations of diabetic retinopathy. After extensive research, we selected a comprehensive dataset comprising 1437 color fundus images, meticulously collected and classified by expert ophthalmologists. Further details about the dataset, including its composition, and the preprocessing steps undertaken to prepare the data for analysis, will be covered in the next subsections. This comprehensive approach ensures that our study is based on reliable and clinically relevant data, enhancing the accuracy and applicability of our findings.

A. Dataset Composition

The dataset "Fundus Images for the Study of Diabetic Retinopathy" [28] comprises 1437 color fundus images acquired at the Department of Ophthalmology, Hospital de

Clínicas, Facultad de Ciencias Médicas, Universidad Nacional de Asunción, Paraguay. Created by a team of researchers and ophthalmologists, this dataset was collected using the Visucam 500 camera from Zeiss, following clinical procedures. Expert ophthalmologists have meticulously classified the images into seven categories: No DR signs (711 images), Mild NPDR (6 images), Moderate NPDR (110 images), Severe NPDR (210 images), Very Severe NPDR (139 images), PDR (116 images), and Advanced PDR (145 images). This classification aids in the detection and study of Non-Proliferative Diabetic Retinopathy (NPDR) and Proliferative Diabetic Retinopathy (PDR) at various stages. The dataset is a valuable resource for researchers and clinicians focusing on the early detection and management of diabetic retinopathy.

B. Dataset Preparation

This section provides a comprehensive overview of the steps undertaken to collect, analyze, and preprocess the data. This includes detailed descriptions of the procedures for data cleaning, formatting, and transformation. We also address any data quality issues encountered during the process and explain how they were resolved. Our primary objective is to ensure that our data preparation process is clearly and transparently documented, supporting the accuracy, reliability, and reproducibility of our research findings. Through this meticulous approach, we aim to establish a robust foundation for our subsequent analysis, ensuring that the data used is of the highest quality and integrity.

1) Data exploration: For this research section, we delved into our database and analyzed the data through diverse charts and summaries. Fig. 1 enabled us to have quantitative measures to evaluate our dataset qualitatively. This visual representation allow us to easily identify trends and patterns within the data, helping us make informed decisions moving forward. By analyzing both the count and percentage of each stage, we can gain a comprehensive understanding of the distribution of diabetic retinopathy stages in our dataset.

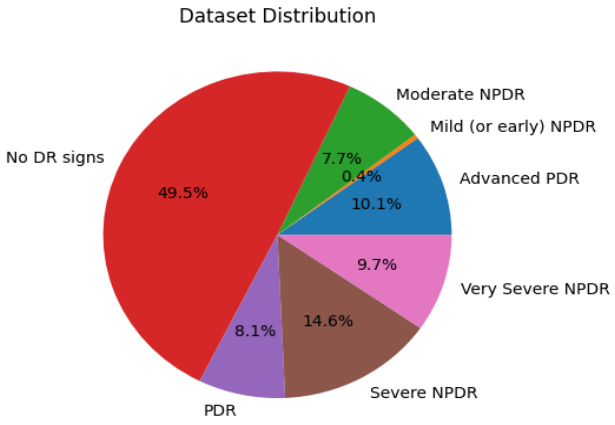


Fig. 1. Percentage distribution of diabetic retinopathy stages.

2) Data preprocessing: Analyzing the distribution graph reveals significant class imbalances. To address the primary stages of interest related to diabetic retinopathy, which include the no diabetic retinopathy (NDR) class, non-proliferative diabetic retinopathy (NPDR), and proliferative diabetic retinopathy (PDR), we merged the pathological sub-stages into these three main categories by assigning each sub-stage to its corresponding main category. This reorganization not only helps balance the distribution in our dataset but also aligns logically with the objectives of our research. The ultimate goal is to develop an automated diagnostic system that assists doctors in accurately predicting the stage of diabetic retinopathy, thereby improving clinical decision-making.

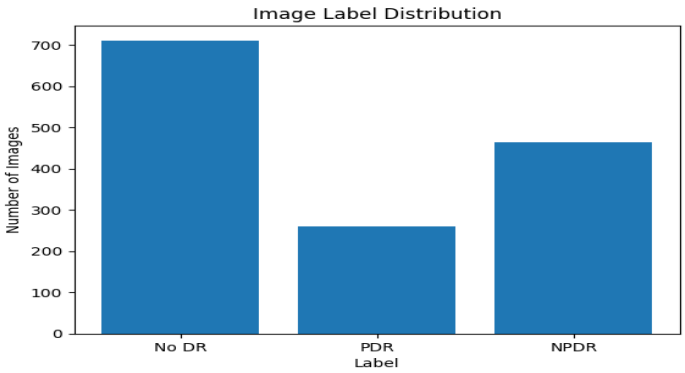


Fig. 2. Post-Merging label distribution of diabetic retinopathy stages.

After analyzing the Fig. 2 representing the distribution of labels post-merging, it is evident that the issue of class imbalance remains prevalent. This imbalance underscores the necessity of implementing data augmentation techniques to create a more balanced dataset, which is crucial for training our models effectively, as it helps prevent biases and improves the overall performance and generalization of the models. Therefore, we must incorporate appropriate data augmentation strategies to address this imbalance and enhance the robustness of our machine-learning models.

To tackle the class imbalance in our dataset, we employed several data augmentation techniques, including horizontal flip, vertical flip, 90-degree rotation, 180-degree rotation, 270-degree rotation, and zoom. These transformations were carefully selected to maintain the realistic characteristics of the original data while increasing the representation of underrepresented classes. By applying these augmentation techniques, we created an enhanced dataset that ensured each class had sufficient samples. The resulting distribution, depicted in Fig. 3, shows a significantly more balanced dataset. This balanced dataset facilitates more accurate and reliable model training, thereby improving performance and generalization.

The dataset was divided into 80% for training and 20% for testing. Subsequently, the training set was further split, allocating 80% for training and 20% for validation. The sample sizes for the three classes (NDR, NPDR, PDR) across the training, validation, and testing sets were as follows: training set - NDR: 1920, NPDR: 1785, PDR: 1001; validation set - NDR: 480, NPDR: 447, PDR: 251; testing set - NDR: 600, NPDR: 558, PDR: 314.

III. Proposed Approach

Developing different pathways to illustrate our models is the main objective of our research which aims at improving the

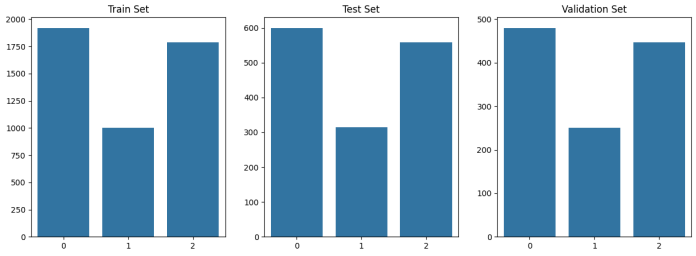


Fig. 3. Distribution of labels in training, testing, and validation sets after data augmentation.

performance and interpretability of predictive models. We will first train and investigate Gradient-weighted Class Activation Mapping (Grad-CAM) technique. It is a powerful visualization tool that allows us to understand how convolutional neural networks (CNNs) make decisions by showing which parts of the input image have the most influence on the model's predictions. After that we will use ensemble methods for enhancing model performance namely meta-learners for increased predictive accuracy. This approach not only reveals interpretative routes of our models but also boosts their prediction abilities significantly.

A. Models Training

1) Elaboration and evaluation of the base CNN model: In our initial training experiment, we designed the convolutional neural network (CNN) depicted in Fig. 4 using Keras to evaluate the model's performance on the retinal image dataset of images resized to 112x112x3. The network architecture included three convolutional layers, each with 64 filters, a 3x3 kernel size, and ReLU activation, followed by 2x2 max-pooling layers, and two fully connected layers with 128 units each, incorporating dropout regularization with a dropout rate of 0.5. The model was compiled with a set of hyperparameters summarized in Table I below and to enhance the effectiveness of our training we used callbacks such as ReduceLROnPlateau to adjust the learning rate by a factor of 0.2 with a patience of 5 epochs (minimum learning rate of 0.0001), EarlyStopping to prevent overfitting with a patience of 10 epochs, and ModelCheckpoint to save the best model based on validation accuracy. Training was initially launched for 80 epochs with validation on a separate test set, providing insights into the model's generalization capabilities and guiding further optimization efforts.

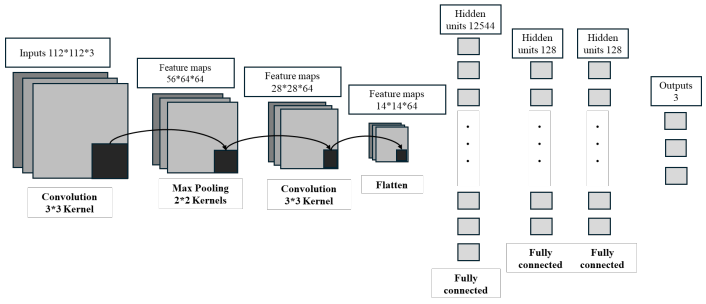


Fig. 4. Layered visualization of base convolutional neural network.

TABLE I. SUMMARY OF HYPERPARAMETERS USED DURING TRAINING

Hyperparameter	Value
Input Size	112×112×3
Batch Size	32
Number of Epochs	80
Learning Rate	0.001
Loss Function	Categorical Crossentropy
Optimizer	Adam

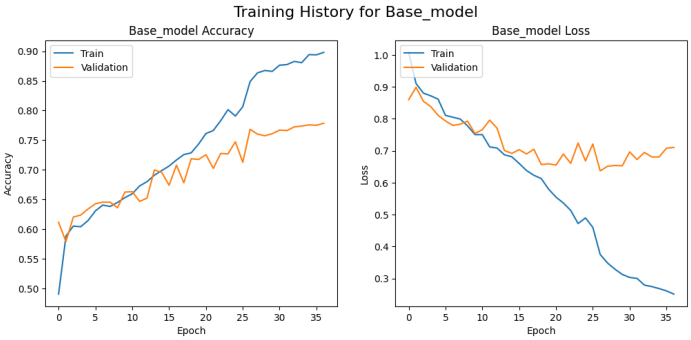


Fig. 5. Training and validation accuracy and loss for base CNN model.

The results of our initial model training are depicted in Fig. 5, illustrating the performance evolution of our base model. The accuracy plot shows a consistent increase in training accuracy, surpassing 90% by the end of the training period, while the validation accuracy exhibits significant fluctuations and stabilizes around 75%. Concurrently, the loss plot indicates a steady decrease in training loss, reaching below 0.3, whereas the validation loss remains relatively high and variable, around 0.7. These observations suggest early signs of overfitting, possibly due to the limited size of our dataset.

To address this issue, we plan to leverage pretrained models for transfer learning. By utilizing a large, pretrained model, we can benefit from its learned representations, which are generally more robust and less prone to overfitting. Transfer learning allows us to adapt these prelearned features to our specific task, typically requiring less data and thereby reducing the risk of overfitting to our training set.

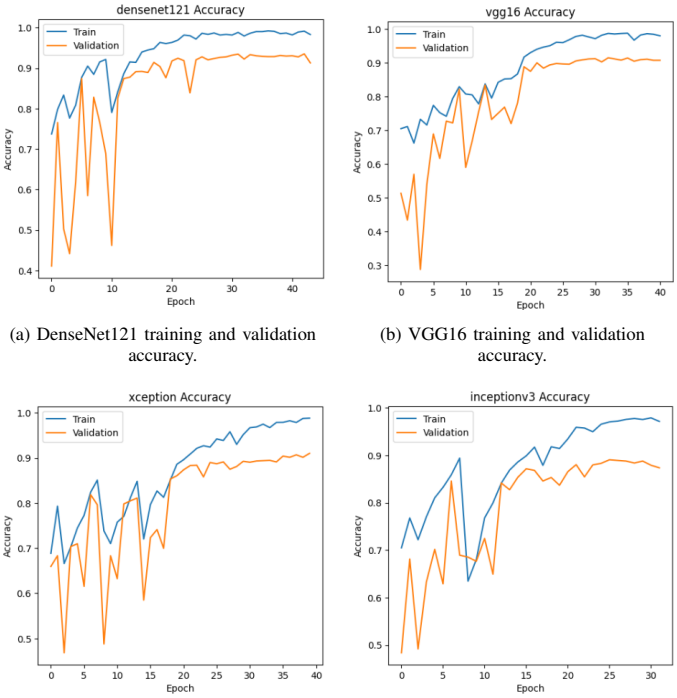
In the upcoming subsection, we will detail our implementation of transfer learning tailored to our task, aiming to maximize the accuracy of our models.

2) Exploring pretrained CNN models for improved performance: In the following section, we considered a selection of pretrained models for our project. The selection of the models was influenced by specific characteristics such as accuracy, architectural diversity, and efficiency as outlined in the table proposed by Keras [30] depicting models performances on the ImageNet validation dataset. Our choice included Xception [31], DenseNet121 [32], VGG16 [33], and InceptionV3 [34] for classifying diabetic retinopathy images, this choice is well-founded due to the high accuracy, architectural diversity, and

efficiency of these models, which are essential for effective classification of diabetic retinopathy images. Xception's depth-wise separable convolutions ensure detailed feature extraction with reduced computational complexity, vital for detecting subtle variations in retinal images. DenseNet121's densely connected layers maximize information flow and feature reuse, enhancing the learning of complex patterns crucial for accurate diagnosis. VGG16's straightforward architecture contributes to robust performance, making it a practical choice for clinical applications due to its ability to deliver consistent results. InceptionV3 achieves a balance between high accuracy and efficiency by utilizing factorized convolutions and dimensionality reduction techniques, which enable effective analysis even with limited computational resources. Together, these models offer a comprehensive approach, combining high performance, diverse methodologies, and operational flexibility, essential for reliable and precise diabetic retinopathy classification.

In our transfer learning experiment, we retrained the four selected models, each adapted for a three-class classification task. The training protocol was meticulously designed to optimize model performance. Each model was initialized with ImageNet weights and fine-tuned by replacing the original classification head with a dense layer of size 3, followed by a softmax activation. The training was conducted using the Adam optimizer with a learning rate of 0.001, minimizing categorical crossentropy loss. The training was initialized for 80 epochs, with the following key hyperparameters that included a batch size of 32, a dropout rate of 0.5 to prevent overfitting, and L2 regularization with a lambda of 0.01. We employed ReduceLROnPlateau to reduce the learning rate by 20% if the validation loss plateaued over 5 epochs (minimum learning rate of 0.0001), EarlyStopping to halt training if validation loss stagnated for 10 epochs, and ModelCheckpoint to save the best-performing model based on validation accuracy. The models were evaluated on a the test validation set, with accuracy and loss monitored throughout the training to assess performance.

The Fig. 6 shows how the four models evolve during the training process on our Diabetic Retinopathy dataset. It illustrates a rapid increase in accuracy in the initial epochs, followed by stabilization at high levels. While the training accuracy of all models rapidly reaches near-perfect levels, the validation accuracy stabilizes slightly lower, typically around 0.85 to 0.90. During the early stages, noticeable fluctuations in both training and validation accuracies are observed, especially within the first 5 to 10 epochs of training. These initial fluctuations can be attributed to the fine-tuning process, where the models, pre-trained on a different task with a different set of images, are adjusting their weights to accommodate the new data. The fluctuations are likely a result of the models attempting to balance learning new features specific to the new dataset while retaining the generalized knowledge acquired from their pre-training. As training continues, the models adapt gradually, resulting in fewer fluctuations and increased stability in accuracy. The results obtained of this operation were as follows, VGG16 demonstrated strong performance, reaching a peak validation accuracy of 91.44% at epoch 33. Despite initial success, the model exhibited some fluctuations in validation accuracy, showing it was affected by changes in the learning rate. However, the overall stability and high accuracy make VGG16 a robust choice for our classification



(a) DenseNet121 training and validation accuracy.

(b) VGG16 training and validation accuracy.

(c) Xception training and validation accuracy.

(d) Inception training and validation accuracy.

Fig. 6. Training and validation evolution of pre-trained models.

task. DenseNet121 showed impressive results, achieving the highest validation accuracy of 93.55% at epoch 43. This model exhibited consistent performance with minor variations in accuracy and loss, reflecting its ability to capture intricate features effectively. InceptionV3 achieved a peak validation accuracy of 92.80% at epoch 38, performing well on the complex data and competing effectively with other models. Xception achieved a peak validation accuracy of 92.35% at epoch 36, showcasing high accuracy and efficient performance, with slight variations in precise validation accuracy and loss metrics.

These results underscore the efficacy of transfer learning in leveraging pretrained models for specialized classification tasks, demonstrating significant potential for diabetic retinopathy classification task.

B. Enhancing Model Interpretability with Grad-CAM

After training models with enhanced accuracy in detecting diabetic retinopathy, it is crucial to implement methods that boost the visibility and interpretability of these deep learning predictions. Such methods enable practitioners to visually interpret model predictions, highlighting areas of interest that may require further examination. Integrating Gradient-weighted Class Activation Mapping (Grad-CAM) with our most precise model, the Densenet model, offers a robust solution.

Grad-CAM is a technique used to produce visual explanations for the decisions made by convolutional neural networks (CNNs). It works by utilizing the gradients of a target concept, such as diabetic retinopathy, flowing into a convolutional

layer of the CNN. By calculating these gradients, Grad-CAM generates a heatmap that shows which regions of the input image are most influential in the model's decision-making process. These heatmaps effectively pinpoint the critical areas within retinal images that the model considers important for diagnosing diabetic retinopathy.

In our approach to generate these Grad-CAM heatmaps, we first constructed a gradient model that outputs the activations from a convolutional layer and the model's predictions. We employed TensorFlow's GradientTape to capture the gradients of the predicted class score with respect to these activations. By computing the mean intensity of the gradients for each output channel of the convolutional layer, we obtained the importance weights. These weights were then applied to the feature maps, resulting in a heatmap that highlights the regions in the image most relevant to the prediction.

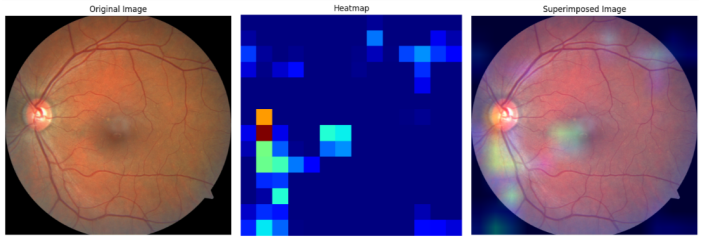


Fig. 7. Grad-CAM visualization for No Diabetic Retinopathy (NDR).

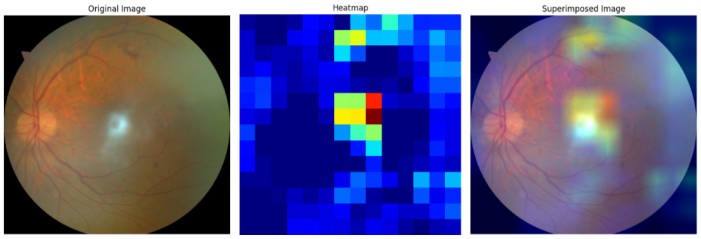


Fig. 8. Grad-CAM visualization for Non-Proliferative Diabetic Retinopathy (NPDR).

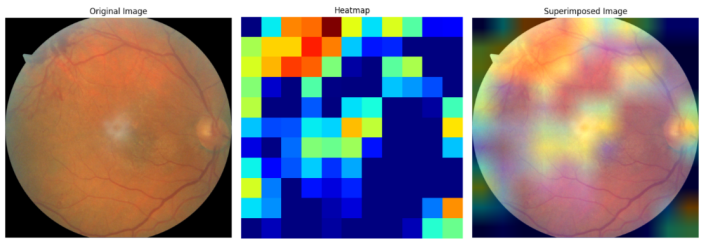


Fig. 9. Grad-CAM visualization for Proliferative Diabetic Retinopathy (PDR).

To visualize the highlighted regions, as depicted in Fig. 7, 8 and 9, which represent the output of our approach for different stages of diabetic retinopathy, we followed a systematic process. First, we loaded the original image and resized the heatmap to match its dimensions. The heatmap was then colored using the jet colormap, scaled appropriately, and superimposed onto the original image with a specified level of transparency. Finally, we displayed the original image, the heatmap, and the superimposed image side-by-side to provide a comprehensive visual analysis of the model's focus areas. This approach allows for a clear and detailed examination of the regions identified by the model, aiding in the interpretation and validation of its predictions.

The three figures illustrate examples of diabetic retinopathy classes: no diabetic retinopathy (NDR), non-proliferative diabetic retinopathy (NPDR), and proliferative diabetic retinopathy (PDR). In each figure, the image on the right shows the original retinograph, commonly used in medical practice. The middle image represents the heatmap, also known as class activation maps (CAM). These maps highlight critical regions in an image responsible for specific predictions made by a convolutional neural network (CNN), obtained by analyzing the flow of gradients in a CNN layer. These maps demonstrate how specific image regions influence the model's predictions.

The Fig. 7 shows a case without diabetic retinopathy, where the model uses the main blood vessels as key features to classify a patient as healthy and without diabetic retinopathy.

The Fig. 8 presents an example of non-proliferative diabetic retinopathy (NPDR), also known as early-stage diabetic retinopathy. This condition results in increased capillary permeability, microaneurysms, hemorrhages, exudates, and complications such as macular ischemia and edema. Our model accurately identifies these features in the three regions of interest, depicted in the left image by the overlay of the original image and the class activation map.

The Fig. 9 illustrates a case of proliferative diabetic retinopathy (PDR), which develops after NPDR and is more severe. It is characterized by the growth of new blood vessels, often accompanied by fibrous tissue growth in front of the retina. These new blood vessels can also form in the front part of the eye, including the iris, contributing to severe vision loss in proliferative retinopathy. These various symptoms are well detected by our model, justifying the extensive regions of interest given the presence of microvascularizations and clear lesions in the example image represented by Fig. 9.

Examining the examples clearly shows the value of using GradCAM to pinpoint specific areas of abnormality, enabling the development of targeted detection and treatment procedures. This advanced technology improves the accuracy and efficiency of detecting retinopathy, leading to better patient outcomes.

C. Boosting Model Performance through Ensemble Methods

In this section, our goal is to enhance the predictive capabilities of our models. To achieve this, we will investigate the collective capabilities of our trained models. Fig. 10 illustrates the conceptual diagram of our proposed approach, leveraging the potential of our four trained models and the generated GRAD-CAM to achieve a balance of precision and interpretability.

Our goal is to enhance prediction accuracy and dependability by leveraging the combined strengths of these models. Fig. 11 illustrates the conceptual diagram of the proposed ensemble learning prediction generation process. By adopting

this strategy, we aim to utilize the variety and complementing qualities of each model, resulting in a stronger and more reliable prediction model.

This approach allows us to make informed decisions based on the insights provided by each model, leading to more robust predictions. Through this ensemble learning technique, we can maximize the potential of our models and improve overall performance.

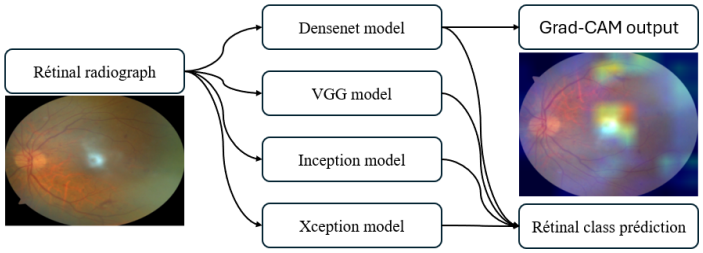


Fig. 10. Conceptual diagram of the proposed approach.

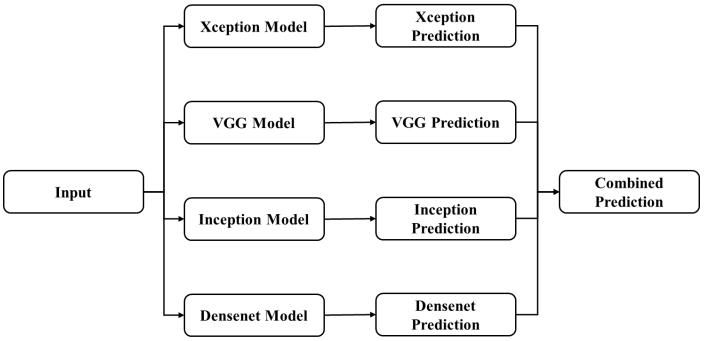


Fig. 11. Conceptual diagram of the proposed ensemble learning prediction.

1) Improving accuracy with weighted ensemble predictions: Within this section, we employed a weighted ensemble of our four trained models to enhance their performance in classifying diabetic retinopathy. The process operates as follows: each model generates predictions on the test samples, and we assess the correctness of each model. Subsequently, these accuracies are utilized to allocate weights to the models, with the more precise models being assigned larger weights. The weights are normalized to ensure that their sum is equal to one.

To get the ultimate forecast for each test sample, I employ a weighted majority voting approach. The predicted result of each model is multiplied by its corresponding weight, and the class with the greatest weighted vote is selected as the ultimate prediction. This approach leverages the advantages of each individual model, leading to a more resilient and precise total prediction.

Mathematically, this can be described as follows:

Let M_i denote the i-th model, and let p_{ij} be the prediction of model M_i for sample j. The accuracy a_i of model M_i is used as its weight w_i, where w_i is normalized so that $\sum_{i=1}^{N} w_i = 1$.

The weighted vote $v_j(k)$ for class k for sample j is given by:

$$v_j(k) = \sum_{i=1}^{N} w_i \cdot I(p_{ij} = k) \tag{1}$$

where, I is the indicator function, which is 1 if $p_{ij} = k$ and 0 otherwise.

The final prediction $\hat{y}_j$ for sample j is:

$$\hat{y}_j = \arg\max_k v_j(k) \tag{2}$$

The following Table II displays the precision of each individual model as well as the collective ensemble model. The findings reveal that the ensemble model obtained a greater accuracy than any of the individual models, proving the usefulness of the weighted ensemble technique.

TABLE II. PERFORMANCE METRICS OF INDIVIDUAL AND ENSEMBLE MODEL

Model	Precision	Recall	F1-Score	Accuracy
Xception Model	0.892542	0.892663	0.892164	0.892663
Vgg Model	0.914305	0.914402	0.914229	0.914402
Inception Model	0.890399	0.890625	0.890181	0.890625
Densenet Model	0.937100	0.935462	0.935564	0.935462
Ensemble (Weighted Voting)	0.939495	0.939538	0.939275	0.939538

By adopting this weighted ensemble technique, the combined model leverages the capabilities of each individual model, leading to increased classification performance, as indicated by the ensemble accuracy of 0.9395.

2) Advanced ensemble technique stacking generalization: Stacking generalization is a sophisticated ensemble learning technique designed to enhance predictive performance by integrating the outputs of numerous base models. Unlike simple averaging or majority voting, stacking involves building a meta-model to learn the best method to combine the predictions of base models. This approach leverages the capabilities of each individual model, leading to more accurate and robust predictions.

The stacking process begins with training multiple base models independently on the training data. These base models can be of various types, or they can be the same type with varying hyperparameters or training sets. Each model is trained to optimize its performance on the given data, creating a diverse set of models with unique strengths and limitations.

Once the base models are trained, the next stage is to generate meta-features. This involves using the trained base models to make predictions on the training set. The predictions from each base model are put together, and if each model outputs a probability for each class, these probabilities are used as meta-features. For instance, in our classification problem with three classes and four base models, each model will output three probabilities per sample, resulting in a total of 12 meta-features per sample (4 models × 3 classes).

The collected meta-features are then used to train a new model, known as the meta-model. The meta-model is trained

on the meta-features generated from the training set, learning how to combine the predictions of the basis models to create the final prediction. Common choices for meta-models include logistic regression, random forests, or another neural network, etc. The meta-model aims to capture intricate relationships between the base model predictions that simple averaging or voting are unable to capture.

Finally, the trained meta-model is used to make final predictions based on the meta-features of the test set. The meta-model processes these meta-features and outputs the final prediction for each sample. This final phase ensures that the strengths of each base model are effectively combined to produce the most accurate predictions.

In our application of this approach, we have leveraged the four models we previously trained and utilized the meta-characteristics generated by these retrained models to train a set of meta-learners. Specifically, we investigated decision trees, multi-layer perceptrons (MLP), Naive Bayes, k-nearest neighbors (KNN), support vector machines (SVM), random forests, and logistic regression. The results obtained from these meta-learners are summarized in the table above. To analyze and determine the best model among those evaluated, we need to consider various factors such as accuracy, precision, recall, and F1-score. Accuracy measures the overall correctness of the model by calculating the ratio of correctly predicted instances to the total instances. Precision is crucial when the cost of false positives is high, as it indicates the proportion of true positive predictions out of all positive predictions made by the model. Recall, also known as sensitivity, is essential in scenarios where false negatives are particularly costly, as it measures the proportion of actual positives that are correctly identified. The F1-score provides a balanced assessment by considering both precision and recall, making it particularly useful when dealing with imbalanced datasets. This comprehensive evaluation will ensure that we identify the most effective model for our specific application.

In order to acquire the findings represented in Table III, we extensively fine-tuned the hyperparameters of each meta-learner. By varying parameters such as learning rate, number of estimators, kernel types, etc. we improved the performance of each model to reach the best potential outcome. The fine-tuning process comprises iterative testing and validation to establish the highest performing meta-models.

For logistic regression, the optimal combination of hyper-parameters was determined to be $C = 10$ and $penalty = l2$, achieving an accuracy of 0.9436. Here, C controls the inverse of the regularization strength, with smaller values indicating stronger regularization, and $penalty$ specifies the norm used in the penalization .

In the case of the random forest, the best performance was obtained with $max\_depth = 20$ and $n\_estimators = 10$, resulting in an accuracy of 0.9341. The max_depth parameter limits the number of levels in each decision tree to prevent overfitting, while $n\_estimators$ defines the number of trees in the forest.

For the support vector machine (SVM), using $C = 10$ and a $kernel = linear$ yielded the highest accuracy at 0.9450. The C parameter is a regularization parameter, and the $kernel$ parameter specifies the kernel type used in the algorithm, with "linear" indicating a linear kernel.

The k-nearest neighbors (KNN) model performed best with $n\_neighbors = 10$ and $weights = uniform$, achieving an accuracy of 0.9429. The $n\_neighbors$ parameter determines the number of neighbors to use, and the $weights$ parameter indicates how the influence of the neighbors is weighted, with "uniform" meaning all neighbors are weighted equally.

The naive Bayes model, which did not require any hyper-parameter tuning, reached an accuracy of 0.9212. Naive Bayes models typically do not have tunable hyperparameters in their basic form.

Lastly, the multi-layer perceptron (MLP) showed optimal performance with $activation = relu$ and $hidden\_layer\_sizes = (100,)$, resulting in an accuracy of 0.9450. The $activation$ parameter specifies the activation function for the hidden layer, with "relu" standing for Rectified Linear Unit, and $hidden\_layer\_sizes$ defines the number of neurons in each hidden layer, with (100,) indicating one hidden layer with 100 neurons.

The decision tree model achieved an accuracy of 0.9307 with $max\_depth = 10$. The max_depth parameter limits the number of levels in the tree, helping to prevent overfitting.

The KNN meta-learner scored the greatest accuracy of 0.9463, with good precision and F1-Scores across all classes, indicating a well-balanced and robust performance. Logistic Regression and Random Forest followed closely with accuracies of 0.9436 and 0.9429, respectively, indicating equal performance in precision, recall, and F1-Score. Naive Bayes and Decision Tree models revealed lower accuracies of 0.9212 and 0.9192, with higher variability in precision between classes.

In summary, stacking is a potent strategy in ensemble learning that may considerably boost prediction performance. By training numerous base models and a meta-model to integrate their outputs, stacking effectively exploits the capabilities of each model as what was proved in our experience. the greatest accuracy of our base models was of 0.9355 and were pushed to 0.9463 performed by the KNN meta-learner resulting to a more precise and reliable predictions wich was the purpose of this investigation .

IV. Experimental Results and Comprehensive Discussion

A. Experimental Results

Our investigation first started by structuring the study area in terms of three classes of important relevance in clinical practice: NDR, NPDR and PDR, followed by the training of a basic convolutional neural network (CNN) using Keras to test its performance on retinal radiographs downsized to 112x112x3. The CNN design featured three convolutional layers with 64 filters, each employing a 3x3 kernel size and Rectified Linear Unit (ReLU) activation functions. This was followed by 2x2 max-pooling layers and two completely linked layers, each with 128 units. We applied dropout regularization with a dropout rate of 0.5 to avoid overfitting. The model was constructed using categorical crossentropy loss as the loss function and employed the Adam optimizer with a learning

TABLE III. PERFORMANCE METRICS OF META-LEARNERS IN ENSEMBLE APPROACH

Meta-Learner	Meta-Model Accuracy	Precision (Class 0)	Precision (Class 1)	Precision (Class 2)	Recall	F1-Score
Logistic Regression	0.943614	0.962046	0.932886	0.929577	0.943614	0.937721
Random Forest	0.942935	0.954323	0.936242	0.934046	0.942935	0.937622
SVM	0.942255	0.958746	0.933333	0.929329	0.942255	0.937170
KNN	0.946332	0.960461	0.937086	0.936057	0.946332	0.941698
Naive Bayes	0.921196	0.962712	0.802778	0.955939	0.921196	0.912088
MLP	0.940897	0.954173	0.935811	0.929204	0.940897	0.935344
Decision Tree	0.919158	0.919094	0.944637	0.906195	0.919158	0.916664

rate set to 0.001. To improve the training process, we implemented callbacks such as ReduceLROnPlateau, EarlyStopping, and ModelCheckpoint. Training ran for 80 epochs, including validation on a unique test set to measure the model's generalization.

The results of our initial model training, depicted in Fig. 5, showed a consistent increase in training accuracy, surpassing 90% by the end of the training period. However, the validation accuracy showed notable fluctuations before stabilizing at around 75%. Concurrently, the training loss steadily decreased, while the validation loss remained relatively high and variable, suggesting early signs of overfitting due to the limited dataset size. To mitigate this, we utilized pretrained models for transfer learning by adapting Xception, VGG16, InceptionV3, and DenseNet121 for a three-class classification task. Each model was initialized with ImageNet weights and fine-tuned by replacing the original classification head with a dense layer of size 3, followed by a softmax activation. Training was conducted using the Adam optimizer with a learning rate of 0.001, a batch size of 32, a dropout rate of 0.5, and L2 regularization with a lambda of 0.01. We employed ReduceLROnPlateau, EarlyStopping, and ModelCheckpoint to optimize the training process.

The training evolutions of these models, depicted in Fig. 6, revealed that VGG16 achieved a peak validation accuracy of 91.44%, DenseNet121 reached the highest validation accuracy of 93.55%, InceptionV3 attained 92.80%, and Xception achieved 92.35%. These results underscore the efficacy of transfer learning in leveraging pretrained models for specialized classification tasks. Subsequently, we integrated Gradient-weighted Class Activation Mapping (Grad-CAM) to improve the interpretability of the models predictions. Grad-CAM generates heatmaps that highlight regions of input images most influential in the model's decision-making process. Fig. 7, 8, and 9 show Grad-CAM outputs for various stages of diabetic retinopathy. These show where the model's focus is and help in understanding and confirming its predictions.

To further enhance the predictive capabilities of our models, we explored ensemble methods. We employed a weighted ensemble of our four trained models, allocating weights based on each model's accuracy. The combined model achieved an accuracy of 93.95%, outperforming the individual models. We explored stacking generalization, an advanced ensemble learning technique that combines the outputs of multiple base models through a meta-model. We trained several meta-learners, including decision trees, multi-layer perceptrons, Naive Bayes, k-nearest neighbors (KNN), support vector machines (SVM), random forests, and logistic regression. The KNN meta-learner achieved the highest accuracy of 94.63%, showcasing superior performance in precision and F1-scores across all classes. Logistic regression and random forest followed closely, with accuracies of 94.36% and 94.29%, respectively.

The workflow outlined in our study, as illustrated by the flowchart in Fig. 12, encapsulates the systematic approach we employed to optimize diabetic retinopathy classification. By combining data preprocessing, transfer learning with multiple deep learning models, and advanced ensemble methods, we were able to progressively enhance model accuracy and robustness. The flowchart also highlights the use of interpretability tools like Grad-CAM, which provided critical insights into the model's decision-making process. This visual representation underscores the complexity and integration of the methodologies discussed, offering a clear, step-by-step view of how each component contributed to the overall success of our approach. This structured methodology not only improved classification performance but also ensured that our models are interpretable and clinically relevant, paving the way for their potential application in real-world settings.

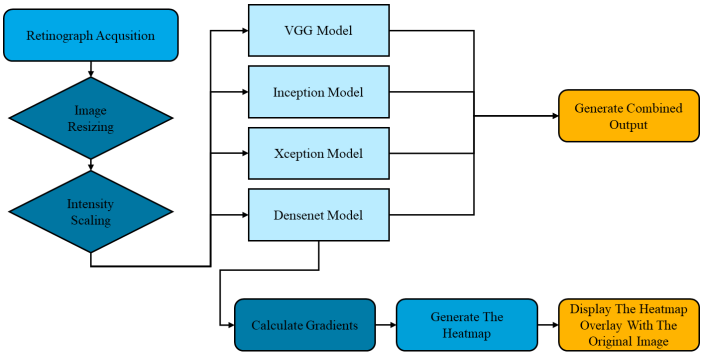


Fig. 12. Workflow for diabetic retinopathy classification using deep learning and ensemble techniques.

B. Discussion

The outcomes of our experiments were the materialization of our conceptual method illustrated in Fig. 10, aiding medical professionals in achieving both precise classification of diabetic retinopathy and visual assistance, thereby providing dual support for clinical decision-making.

Our experiments yielded several important discoveries that can be summarized as follows:

1) Transfer learning effectiveness: Pretrained models, particularly DenseNet121, substantially enhanced classification performance compared to the initial base model. This demonstrates the value of utilizing pretrained networks.

2) Enhanced performance through ensemble methods: Both weighted ensemble and stacking generalization techniques effectively boosted the predictive accuracy of our models. Notably, the KNN meta-learner achieved the highest performance, showcasing the power of combining multiple models to capture diverse patterns in the data.

3) Interpretability through Grad-CAM: The Grad-CAM visualizations provided meaningful insights into the decision-making process of the models, enhancing the interpretability and trustworthiness of the predictions.

V. Conclusion and Perspectives

In conclusion, our comprehensive approach integrated transfer learning, Grad-CAM for interpretability, and ensemble methods, resulting in significant improvements in the performance and reliability of our predictive models. Testing demonstrated that pre-trained models, advanced visualization techniques, and sophisticated ensemble strategies markedly enhance deep learning models for classifying diabetic retinopathy. This methodology not only improved model accuracy but also strengthen reliability.

Overall, our approach underscored the importance of leveraging diverse deep learning techniques to elevate the performance of predictive models in medical image classification. By incorporating these methods, we achieved substantial gains in accuracy and reliability for detecting diabetic retinopathy. The combination of these techniques not only enhances their applicability in identifying abnormalities but also paves the way for discovering new diseases and developing treatment strategies. Advanced visualization techniques empower medical professionals to visually cluster abnormalities, ensuring the development of robust and dependable systems that minimize flaws inherent in automated methods.

References

[1] M. Z. Banday, A. S. Sameer, and S. Nissar, *Pathophysiology of diabetes: An overview*, Avicenna Journal of Medicine, vol. 10, no. 4, pp. 174-188, Oct. 2020, doi: 10.4103/ajm.ajm_53_20.

[2] World Health Organization, *Diabetes*, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/diabetes. [Accessed: Jul. 5, 2024].

[3] P. Soewondo, A. Ferrario, and D. L. Tahapary, *Challenges in diabetes management in Indonesia: a literature review*, Global Health, vol. 9, p. 63, Dec. 2013, doi: 10.1186/1744-8603-9-63.

[4] National Eye Institute, *Diabetic Retinopathy*, 2023. [Online]. Available: https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/diabetic-retinopathy. [Accessed: Jul. 6, 2024].

[5] O. Manchadi, F. E. Ben-Bouazza, and B. Jioudi, *Predictive Maintenance in Healthcare System: A Survey*, IEEE Access, vol. 11, pp. 61313-61330, 2023, doi: 10.1109/ACCESS.2023.3287490.

[6] O. Manchadi, F. E. Ben-Bouazza, Z. El Otmani Dehbi, A. Edder, I. Tafala, M. Et-Taoussi, and B. Jioudi, *An Internet of Things-based Predictive Maintenance Architecture for Intensive Care Unit Ventilators*, International Journal of Advanced Computer Science and Applications, vol. 15, no. 2, 2024, doi: 10.14569/IJACSA.2024.0150294.

[7] W. M. Gondal, J. M. Köhler, R. Grzeszick, G. A. Fink, and M. Hirsch, *Weakly-supervised localization of diabetic retinopathy lesions in retinal fundus images*, in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, pp. 2069–2073, Sep. 2017.

[8] Z. Wang, Y. Yin, J. Shi, W. Fang, H. Li, and X. Wang, *Zoom-in-net: Deep mining lesions for diabetic retinopathy detection*, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 267–275, Springer, Berlin, Germany, 2017.

[9] G. Quellec, K. Charrière, Y. Boudi, B. Cochener, and M. Lamard, *Deep image mining for diabetic retinopathy screening*, Medical Image Analysis, vol. 39, pp. 178–193, Jul. 2017.

[10] T. Chandrakumar and R. Kathirvel, *Classifying diabetic retinopathy using deep learning architecture*, International Journal of Engineering Research and Technology, vol. 5, no. 6, pp. 19–24, Jun. 2016.

[11] W. R. Memon, B. Lal, and A. A. Sahto, *Diabetic retinopathy*, The Professional Medical Journal, vol. 24, no. 2, pp. 234–238, 2017.

[12] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, *Convolutional neural networks for diabetic retinopathy*, Procedia Computer Science, vol. 90, pp. 200–205, Dec. 2016.

[13] D. Yang, X. Lin, Y. Gao, and X. Wu, *Lesion detection for diabetic retinopathy based on convolutional neural networks*, in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 2630–2634, Sep. 2017.

[14] P. Garcia, C. Rodriguez, and S. Martinez, *Detection of diabetic retinopathy using a convolutional neural network*, Journal of Computer and Communications, vol. 5, pp. 1–7, 2017.

[15] S. Dutta, B. C. Manideep, S. M. Basha, R. D. Caytiles, and N. C. S. N. Iyengar, *Classification of diabetic retinopathy images by using deep learning models*, International Journal of Grid and Distributed Computing, vol. 11, no. 1, pp. 89–106, Jan. 2018.

[16] J. Luo, H. Zhang, and Y. Li, *Multi-View Diabetic Retinopathy Detection: A Method Combining DCNNs and Attention Mechanisms*, IEEE Transactions on Medical Imaging, vol. 40, no. 2, pp. 407–418, 2021.

[17] L. Chen, *Early recognition of diabetic retinopathy using a multi-scale shallow convolutional neural network*, Journal of Healthcare Engineering, vol. 2020, pp. 1–9, 2020.

[18] F. J. Martinez-Murcia and J. Ortuno, *Regular CNN for Routine Diabetic Retinopathy Diagnosis: Challenges and Solutions*, Computers in Biology and Medicine, vol. 131, p. 104245, 2021.

[19] P. Deepa, G. Selvaraj, and R. T. Selvi, *Efficient Deep Convolutional Neural Network (MPDCNN) for Fundus Image Recognition*, Journal of Digital Imaging, vol. 35, pp. 312–325, 2022.

[20] A. Das, K. Roy, and K. Biswas, *Segmented fundus images for the detection of diabetic retinopathy using a deep learning architecture*, Journal of Digital Imaging, vol. 34, pp. 735–746, 2021.

[21] D. Kalyani and G. Rao, *A reformed capsule network for feature extraction from fundus images*, IEEE Transactions on Medical Imaging, vol. 42, no. 4, pp. 1032–1043, 2023.

[22] S. Oh, J. Kim, and K. Lee, *Fundus photography and deep learning for diabetic retinopathy detection*, IEEE Transactions on Medical Imaging, vol. 40, no. 10, pp. 2472–2482, 2021.

[23] K. Erciyas and N. Barısci, *ROI extraction and deep learning techniques for automatic lesion detection*, Journal of Digital Imaging, vol. 34, pp. 1236–1245, 2021.

[24] S. Azeroual, F. E. Ben-Bouazza, A. Naqi, and R. Sebihi, *Predicting disease recurrence in breast cancer patients using machine learning models with clinical and radiomic characteristics: a retrospective study*, J Egypt Natl Canc Inst, vol. 36, no. 1, p. 20, Jun. 2024, doi: 10.1186/s43046-024-00222-6.

[25] I. Tafala, F.-E. Ben-Bouazza, O. Manchadi, M. Et-Taoussi, and B. Jioudi, *Cephalometric Landmarks Identification Through an Object Detection-based Deep Learning Model*. [Online]. Available: https://api.semanticscholar.org/CorpusID:268418674.

[26] A. Edder, F.-E. Ben-Bouazza, and B. Jioudi, *SkinNet: Enhancing Dermatological Diagnosis Through a New Deep Learning Framework*, in *International Conference on Advanced Intelligent Systems for Sustainable Development (AI2SD'2023)*, M. Ezziyyani, J. Kacprzyk, and V. E. Balas, Eds., Springer Nature Switzerland, Cham, pp. 173–188, 2024, doi: 10.1007/978-3-031-52388-5-17.

[27] Y. Guex-Crosier and F. Behar-Cohen, *Ophtalmologie: Rétinopathie diabétique : nouvelles possibilités thérapeutiques*, Rev Med Suisse, vol. 11, no. 45657, pp. 101–107, 2015, doi: 10.53738/REVMED.2015.11.456-57.0101.

[28] V. E. Castillo Benítez, I. Castro Matto, J. C. Mello Román, J. L. Vázquez Noguera, M. García-Torres, J. Ayala, D. P. Pinto-Roa, P. E. Gardel-Sotomayor, J. Facon, and S. A. Grillo, *Dataset from fundus images for the study of diabetic retinopathy*, Data in Brief, vol. 36, p. 107068, 2021.

[29] V. E. Castillo Benítez, I. Castro Matto, J. C. Mello Román, J. L. Vázquez Noguera, M. García-Torres, J. Ayala, D. P. Pinto-Roa, P. E. Gardel-Sotomayor, J. Facon, and S. A. Grillo, *Dataset from fundus images for the study of diabetic retinopathy*, Zenodo, 2021, doi: 10.5281/zenodo.4647952.

[30] F. Chollet *et al.*, *Keras Applications API*, 2023. [Online]. Available: https://keras.io/api/applications/. [Accessed: May 22, 2024].

[31] F. Chollet, *Xception: Deep Learning with Depthwise Separable Convolutions*, 2017. [Online]. Available: https://arxiv.org/abs/1610.02357.

[32] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, *Densely Connected Convolutional Networks*, 2018. [Online]. Available: https://arxiv.org/abs/1608.06993.

[33] K. Simonyan and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*, 2015. [Online]. Available: https://arxiv.org/abs/1409.1556.

[34] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, *Rethinking the Inception Architecture for Computer Vision*, 2015. [Online]. Available: https://arxiv.org/abs/1512.00567.

Breast Tumor Classification Using Dynamic Ultrasound Sequence Pooling and Deep Transformer Features

Mohamed A Hassanien[1], Vivek Kumar Singh[2], Mohamed Abdel-Nasser[3], Domenec Puig[4]
Department of Computer Engineering and Mathematics, University Rovira i Virgili, Tarragona, Spain 43007[1,4]
Barts Cancer Institute (BCI), Queen Mary University of London, London, UK[2]
Electrical Engineering Department-Faculty of Engineering, Aswan University, Aswan, Egypt[3]

***Abstract*—Breast ultrasound (BUS) imaging is widely utilized for detecting breast cancer, one of the most life-threatening cancers affecting women. Computer-aided diagnosis (CAD) systems can assist radiologists in diagnosing breast cancer; however, the performance of these systems can be degrade by speckle noise, artifacts, and low contrast in BUS images. In this paper, we propose a novel method for breast tumor classification based on the dynamic pooling of BUS sequences. Specifically, we introduce a weighted dynamic pooling approach that models the temporal evolution of breast tissues in BUS sequences, thereby reducing the impact of noise and artifacts. The dynamic pooling weights are determined using image quality metrics such as blurriness and brightness. The pooled BUS sequence is then input into an efficient hybrid vision transformer-CNN network, which is trained to classify breast tumors as benign or malignant. Extensive experiments and comparisons on BUS sequences demonstrate the effectiveness of the proposed method, achieving an accuracy of 93.78%, and outperforming existing methods. The proposed method has the potential to enhance breast cancer diagnosis and contribute to lowering the mortality rate.**

***Keywords*—Breast ultrasound; breast cancer; CAD systems; deep learning; vision transformer**

I. Introduction

Breast cancer in women is one of the most life-threatening cancers worldwide [1], [2]. Early detection significantly reduces the mortality rate, and mammography, an X-ray imaging technique of the breast, remains the gold standard for population-based breast cancer screening. Despite its effectiveness in detecting breast cancer, mammography has limitations, including low sensitivity and high false-positive rates, where normal cases are incorrectly classified as cancerous. To address these limitations, alternative imaging technologies such as magnetic resonance imaging (MRI), 3D tomosynthesis, and ultrasound are often used [3].

Breast ultrasound (BUS) imaging has been effectively used in the detection and diagnosis of breast cancer, specially in the case of women having dense breast tissue or with cases who are at high risk of developing breast cancer [4], [5]. The main merits of BUS imaging are non-invasive and non-ionizing technology, widely available and cost-effective solution, and capable of producing real-time images, which can enhance breast cancer detection sensitivity.

In the last two decades, various computer-aided detection (CAD) systems have been developed for breast cancer detection. In particular, artificial intelligence (AI) based CAD systems have helped in detecting breast cancer early by assisting radiologists in interpreting medical images, including BUS images [6]. AI-powered CAD systems can analyze images quickly and accurately, detecting subtle abnormalities and highlighting region of interest (ROIs), thereby boosting the sensitivity and specificity of breast cancer detection. It should be noted that BUS imaging has some limitations, notably poor contrast, speckle noise, and shadowing artifacts, which can degrade image quality and complicate interpretation (see Fig. 1). These issues make it challenging to differentiate between various tissues and structures. Additionally, BUS is operator-dependent, with image quality varying based on the skill and experience of the sonographer. This highlights the need for effective image processing, noise mitigation techniques, and robust AI-based image classification models, to improve the performance of these CAD systems in breast cancer detection.

In recent years, deep learning has significantly enhanced the automated analysis of BUS images over the past decade by extracting powerful representations from them. This has led to the development of several deep-learning aid (DLA) tools for detecting breast cancer and distinguishing between benign and malignant tumors. Recently, several DLA tools has been proposed, for instance, Ellis et al. of [7] explored deep learning as a classification tool for detecting cancerous ultrasound breast images, aiming to develop a simple, mobile-based classifier. With ResNet50, the CAD system achieved an accuracy of approximately 64% with minimal images, and up to 78% when pretrained. The authors of [8] introduced a novel few-shot learning approach for classifying ultrasound breast cancer images, leveraging the power of meta-learning techniques. Specifically, the authors employed prototypical networks and model-agnostic meta-learning (MAML) algorithms to enable our model to learn from limited data and adapt to new, unseen breast cancer images. Lanjewar et al. [9] integrated three widely used pretrained Convolutional Neural Network (CNN) models, namely, MobileNetV2, ResNet-50, and VGG16 with a long short term memory (LSTM) to extract features from BUS images. The authors used the synthetic minority over-sampling with Tomek (SMOTETomek) method in order to balance the number of extracted features. With the VGG16 model, they achieved an F1 score of 99.0%, Kappa coefficient of 98.9%, and an area under the curve (AUC) of 1.0.

The majority of existing studies have focused on classifying breast tumors using only one ultrasound image per

tumor, whereas some studies, such as [10], [11], have utilized BUS sequences for detecting breast cancer malignancy. In this paper, we consider the quality of BUS images when designing the classification model. In particular, we propose a novel approach for breast tumor classification utilizing dynamic pooling of BUS image sequences. Specifically, this new method captures the temporal evolution of breast tissues in BUS sequences, effectively mitigating the influence of noise and artifacts. The dynamic pooling weights are computed based on image quality metrics, including blurriness and brightness. The resulting pooled BUS sequence is processed by the MobileViTv3 network, which is trained to classify breast tumors as either benign or malignant.

The remainder of this research is organized as follows: Section II reviews related work on breast lesion classification in ultrasound images. Section III details the proposed method. Section IV presents and discusses the experimental results. Finally, Section V concludes the study and provides suggestions for future work.

II. Related Work

It should be noted that most existing breast cancer CAD systems are trained to receive one ultrasound image (OUI) to determine whether it is benign or malignant [12], [13], [14], [15]. For instance, He et al. [16] proposed a new method for breast cancer classification using a wavelet-based vision transformer network. By incorporating the discrete wavelet transform (DWT) into the network input, we enhance the neural network's receptive fields, enabling the capture of significant features in the frequency domain. The proposed model effectively captures intricate characteristics of breast tissue, allowing for accurate breast cancer classification with high precision and efficiency. We evaluated the model using two breast tumor ultrasound datasets, comprising 780 cases from Baheya hospital in Egypt and 267 patients from the UDIAT Diagnostic Centre of Sabadell in Spain. The results show that the proposed transformer network achieves outstanding performance in breast cancer classification, with an AUC scores of 0.984 and 0.968 on both datasets.

Some recent studies showed that BUS sequences may give better detection results. For instance, the authors of [11] proposed a four-stage CAD system: super-resolution calculation, ROI extraction, feature extraction, and classification. The authors used five manually designed features, derived from various image analysis techniques, including GLCM, LBP, HOG, phase congruency-based LBP, and pattern lacunarity spectrum, to classify breast tumors into malignant and benign categories from a BUS image. However, this conventional approach has several limitations, including being computationally time-consuming, less resilient, and requiring specific feature choices and preprocessing activities.

To handle this issues mentioned above, recent studies employed deep learning networks for feature extraction. For instance, Yang et al. [17] presented a temporal sequence dual-branch network (TSDBN) breast cancer classification based on BUS and contrast-enhanced ultrasound (CEUS) sequences. It has two branches: one for BUS sequences and the other for CEUS sequences. In the branch of the BUS sequences, the ResNeXt-18 is employed. In the other branch, temporal sequence regression and a shuffle temporal sequence mechanisms are employed to enhance the temporal features of CEUS sequnces. They used a private dataset to evaluate their method. The dataset has 268 samples: 146 malignant and 122 benign. For each case, the BUS and CEUS sequences were recorded. TSDBN achieved an accuracy of 92.2%. One of the main limitation of this method is that it requires US and CEUS sequences for the same cases, which may not be available. Also, it does not consider the effect of noise and artifacts in US and CEUS sequences in the classification results. The study of [18] used 3D ResNet-50 to classify breast lesions in BUS sequences and 2D ResNet-50 to classify the same lesions in static images, finding that the BUS sequences lead to a higher AUC value of 0.969.

Han et al. [19] presented a ResViT model that combines residual neural networks with vision transformer to extract features from CEUS sequences, and employed a temporal segment network (TSN) to aggregate the spatio-temporal features of all frames in the input sequences. They achieved an accuracy of 78.79% with a private CEUS sequence dataset. Zhang et al. [20] proposed a segment-attention generator (SAG) module that can help deep learning classification models to focus on BUS sequence segments that have clear appearances for classifying breast lesions. The study of [10] introduced a deep-learning-based radiomics approach utilizing BUS sequences, comprising three key components. The ConvNeXt network, a deep CNN trained in the vision transformer style, is employed for radiomic feature extraction. An efficient pooling mechanism is also proposed to combine the malignancy scores of each breast US sequence frame, based on image-quality statistics. Finally, visual interpretations are provided to facilitate understanding. However, this methods achieved acceptable results, there is a big room for further enhancing the classification accuracy.

As mentioned earlier, a common limitation of existing studies is that they neglect the temporal information and image quality of BUS videos when developing classification models. Moreover, they rely on a single BUS image to develop their methods. However, the noisy nature of BUS images, the similarity between normal and abnormal tissues, and the degradation of image quality due to dense breast fat and glandular tissue, which attenuate ultrasonic waves, make accurate diagnosis challenging. These issues pose a significant challenge to building a robust BUS image classification model. The proposed method consider that the temporal information embedded in the BUS sequences and mitigate the effect of the noise utilizing the the weighted dynamic pooling technique.

III. Material and Method

A. Dataset

The proposed CAD system was developed and evaluated using a database consisting of 31 malignant and 28 benign BUS sequences, with each sequence corresponding to a single patient. The BUS sequences created by the Engineering Department of Cambridge University[1]. This dataset is a subset of a larger clinical database of ultrasonic radiofrequency strain imaging data, which was created by the Engineering Department at Cambridge University. The dataset includes 3911

[1] http://mi.eng.cam.ac.uk/research/projects/elasprj/

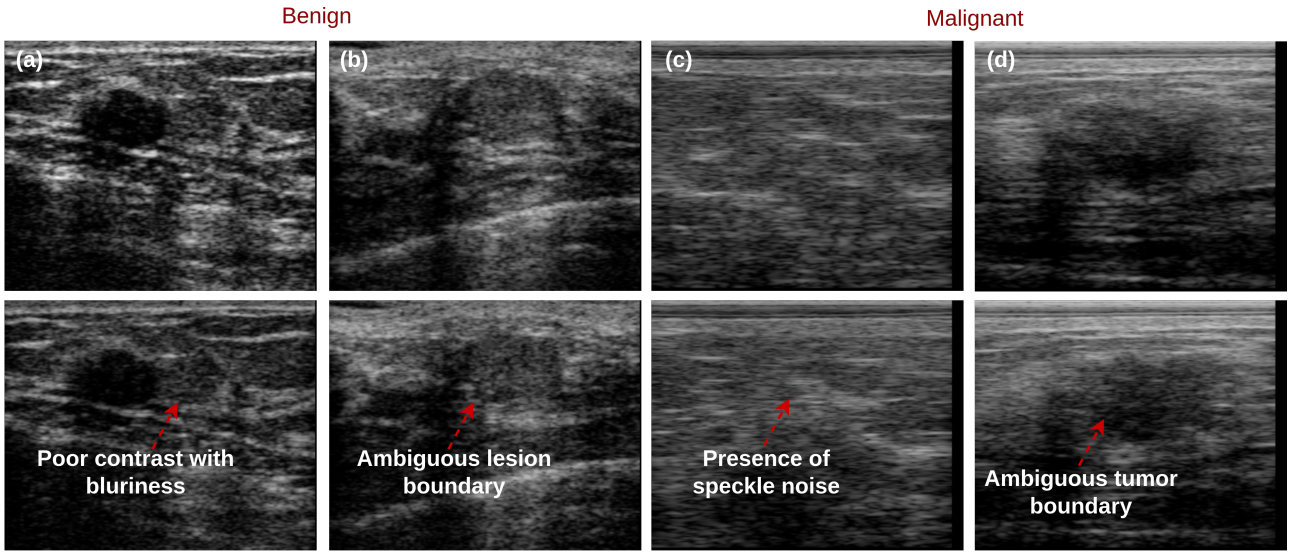


Fig. 1. Examples of BUS images having malignant tumors and benign lesions with various artifacts and challenges like poor contrast, ambiguous lesion or tumor boundaries, and speckle noise.

images containing benign tumors and 5245 images having malignant tumors.

B. Proposed Method

Fig. 2 presents the key components of the proposed method: 1) generating dynamic BUS image from the input BUS sequence, 2) applying transfer learning on MobileViTv3 [21] to extract local and global features from BUS images to differentiate between benign and malignant tumors, and 3) employing different visual interpretation methods to explain the decisions of the classification model.

1) Dynamic BUS sequence pooling: Let $S = [s_1, s_2, \ldots, s_M]$ is an input BUS sequence, where s_i

$$q_t = \frac{1}{W}\sum_{i=t}^{t+W} s_i. \tag{1}$$

where W is a time window.

TVA can be expressed as follows:

$$q_t = \frac{\frac{1}{t}\sum_{i=1}^{t} s_i}{\|\frac{1}{t}\sum_{i=1}^{t} s_i\|}. \tag{2}$$

After obtaining the smoothed version of BUS images, a rank-pooling method can be employed to learn the relative ranks of the BUS images in the input sequence, for instance q_n comes after q_{n-1}, q_{n-1} comes after q_{n-2}, and so on. This relative ranks can expressed as follows:

$$q_n \succ q_{n-1} \succ q_{n-2} \ldots q_1 \tag{3}$$

The rank-pooling technique is used to learn pairwise linear functions $\rho(q_t;\alpha)$, where $\alpha \in R^D$. The ranking score of q_t is computed as $\rho(q_t;\alpha) = \alpha^T.q_t$.

The parameters α of $\rho(q_t;\alpha)$ are optimized using the following objective function [22]:

$$\arg\min_{\alpha} \quad \frac{1}{2}\|\alpha\|^2 + \delta \sum_{\forall i,j,q_i \succ q_j} \theta_{ij}, \quad s.t. \quad \alpha^T(q_{t_i} - q_{t_j}) \geq 1 - \theta_{ij}, \quad \theta_{ij} \geq 0, \tag{4}$$

In this expression, δ stands for the regularization parameter and θ stands for the tolerance margin. The term $\{\alpha^T(q_{t_i} - q_{t_j}) \geq 1 - \theta_{ij}\}$ represents the the constraint of the objective function $\forall t_i, t_j \quad q_{t_i} \succ q_{t_j} \iff \alpha^T.q_{t_i} \succ \alpha^T.q_{t_j}$.

2) Image quality-aware dynamic BUS image generation: The quality of the BUS images can considered when generating the dynamic BUS image from the input BUS sequence by modifying Eq. 1 and Eq. 2 as follows:

$$q_t = \frac{1}{W}\sum_{i=t}^{t+W} \omega_i.s_i. \tag{5}$$

$$q_t = \frac{\frac{1}{t}\sum_{i=1}^{t} \omega_i.s_i}{\|\frac{1}{t}\sum_{i=1}^{t} \omega_i.s_i\|}. \tag{6}$$

where ω_i represents the quality of the th BUS image in the input sequence. The value of w_i may be 0 or 1, where a value of 1 denotes that the BUS image quality exceeds the thresholds of the BUS image quality metrics.

In this study, two efficient general-purpose image quality assessment metrics are used to estimate the quality of BUS images, namely the brightness and blurriness [23], [24], [25].

BUS image blurriness metric: Here, we employ the image blurriness measure presented in [24], where a Gaussian filter

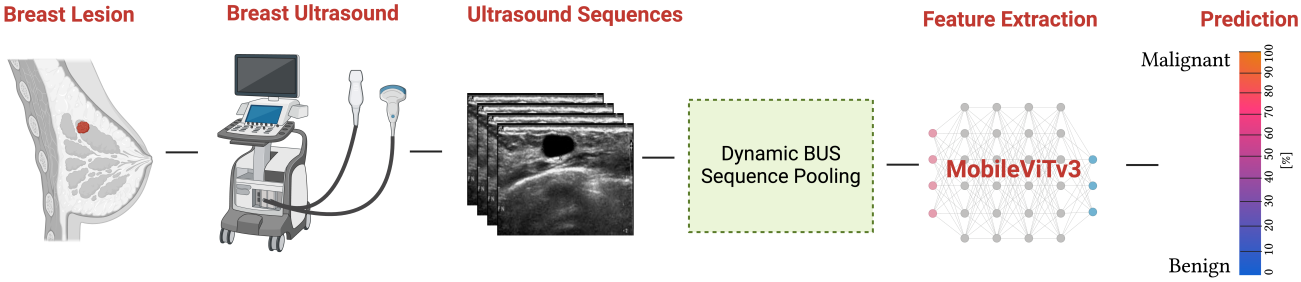


Fig. 2. Overview of the proposed method which consists of three consecutive steps: dynamic BUS sequence pooling, feature extracting using MobileViTv3 and malignancy classification.

is used to suppress the noise from the input image. Let I is a BUS image, the Gaussian filter can be written as:

$$f(a,b) = \frac{1}{(2\pi\sigma^2)} e^{-\frac{\left(a^2+b^2\right)}{2\sigma^2}}, \tag{7}$$

Here, σ is the standard deviation of the Gaussian distribution, and a and b are the coordinates of I.

After suppressing the noise, the variance of Laplacian operator is computed and used as a blurriness score. The 2D Laplacian operator can be expressed as follows:

$$\nabla^2 I(a,b) = \frac{\partial^2 I}{\partial a^2} + \frac{\partial^2 I}{\partial b^2}, \tag{8}$$

BUS images with a blurriness score lower than a threshold are considered as blurry images. Following [24], the blurriness threshold is set to the average blurriness value of benign and malignant BUS images from the training dataset.

BUS image brightness/darkness metric: In this study, we employ the brightness estimation algorithm proposed by Bezryadin et al. [25] as a BUS image quality metric. Following the study of [10], we selected the range from 10 to 30 for the brightness score.

3) Feature extraction: The main powerful approaches to extract features from images are CNNs and ViTs. In the context of breast tumors in ultrasound images, previous studies such as [9], [7] focused in the use of CNN models, [26], [27] used vision transformers, while [28] combined the decisions of different CNN and vision transformers. ViTs produce features representing global information in the images, due to their self-attention mechanism. CNNs extract local features in the images. Several hybrid models have emerged, integrating the strengths of both CNNs and ViTs into a single architecture. By combining the self-attention mechanism of ViTs, which excels at capturing long-range dependencies, with the local kernels of CNNs, which are adept at extracting local information, these models aim to achieve superior performance on various vision tasks. In order to extract robust descriptors from BUS images, in this paper we employ one of the most effective deep learning model that combines CNNs and ViTs, namely, MobileViTv3 [21].

Fig. 3 shows the block diagram of the MobileViTv3 that contains three blocks: local representation (LR) block, global representation (GR) block, and fusion block. The LR block (CNN components) consists of two layers: a 3×3 depthwise convolution layer and a 1×1 convolution layer. The GR block (ViT components) includes N linear transformations (self-attention). The fusion block uses 1×1 convolution layer to fuse the local and global features.

This study involves adapting and training various self-attention based deep vision transformer architectures to extract robust features, which can classify breast cancers as benign or malignant and predict the malignancy score of each input ultrasound image. By leveraging the transfer learning theory, the pre-trained vision transformer network and its parameters can be fine-tuned and applied to the target breast ultrasound dataset, enabling effective knowledge transfer. A support vector machine (SVM) classifier with a radial basis function (RBF) is used for classification.

4) Visual interpretation: To produce visual interpretations (explanations) for the proposed breast tumor classification model, this study employs the Grad-CAM (Gradient-weighted Class Activation Mapping) [29] and Local Interpretable Model-agnostic Explanations (LIME) techniques [30]. Let's denote the input image as I, the class of interest (benign or malignant) as c, and the output probability of the class as $P(c|I)$. The goal of Grad-CAM is to generate a heatmap $L^c_{Grad-CAM}$ that highlights the important regions of each BUS image that contribute to the prediction of the proposed model.

LIME is a model-agnostic method, which works by generating a dataset of similar instances around a specific instance for which we want to understand the model's prediction. Let's denote the original machine learning model as f, and the instance for which we want to explain the prediction as $\mathbf{x}$. LIME generates a dataset of m perturbed instances around $\mathbf{x}$, denoted as $\mathbf{x}'$, by randomly sampling from a distribution $\pi(\mathbf{x}'|\mathbf{x})$. The perturbed instances $\mathbf{x}'$ are then used to generate a new dataset $\mathcal{D} = (\mathbf{x}'_1, f(\mathbf{x}'_1)), \ldots, (\mathbf{x}'_m, f(\mathbf{x}'_m))$. Next, LIME trains an interpretable model g (e.g. a linear model) on the dataset $\mathcal{D}$ to approximate the behavior of the original model f locally around $\mathbf{x}$.

C. Evaluation metrics

To assess the performance of the proposed method for breast tumor classification in BUS sequences, we employ four well-known evaluation metrics, namely, the accuracy, precision, recall, and F1-score. The mathematical expression

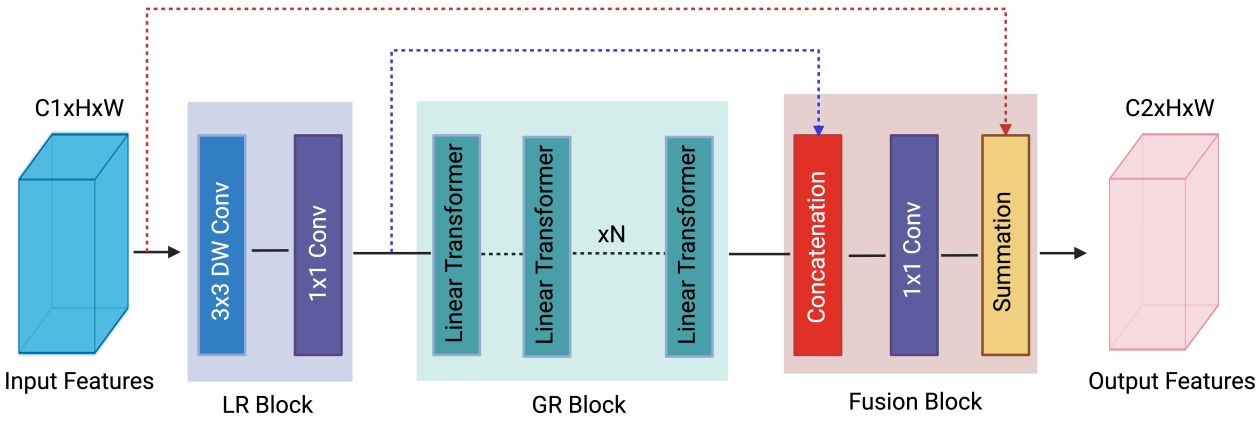


Fig. 3. Overview of the block diagram of the MobileViTv3 that consists of three interconnected blocks: LR, GR, and fusion.

of each evaluation metric is given below:

$$\text{Accuracy} = \frac{TP + TN}{P + N} \tag{9}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{10}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{11}$$

$$\text{F1-score} = \frac{TP}{TP + 0.5(FP + FN)} \tag{12}$$

In these expressions, TP and TN denote the number of BUS sequences having malignant and benign tumors that have been successfully detected by the proposed method, respectively. Conversely, FN stands for the number of malignant tumors wrongly identified by the proposed method as benign tumors. FP stands for the number of benign tumors wrongly identified by the proposed method as malignant tumors.

IV. Experimental Results

A. Training Details

As the BUS images varied in size, all were resized to 224×224 pixels. The AdamW optimizer was employed with an initial learning rate of 0.001, a weight decay of 0.01, and a cosine learning rate scheduler, using binary cross-entropy loss to optimize the model. The training process was conducted with a batch size of two images over 50 epochs. To augment the training data, the input BUS sequences were split into overlapping sub-sequences with a window size of 20, generating multiple dynamic BUS images from each sub-sequence. Additional data augmentation techniques were applied, including 90-degree image rotation, 0.2 image scaling, horizontal flipping (with a probability of 0.5), median filter blurring, and contrast-limited adaptive histogram equalization. All models were developed in Python using the PyTorch framework and trained on an NVIDIA GeForce GTX 1070Ti GPU with 8 GB of RAM.

B. Results

Table I compares various backbone feature extractors, including CNNs, vision transformers, and hybrid models. MobileViTv3-S emerges as the top-performing model, significantly outperforming others with an accuracy of 89.33%, which is more than 1.2% higher than the second-best model, ConvNeXt V2. This demonstrates the strength of hybrid architectures like MobileViT, which combine the local feature-capturing ability of CNNs with the global context awareness of transformers. Despite its depth, ResNet-150 performs good but does not achieved similar results as MobileViTv3-S, recommending that deeper CNNs do not necessarily yield better performance in this task.

The transformer-based models, DEiT and BEiT v2, show satisfactory performance, indicating that transformers may require more fine-tuning for optimal results. XCiT performs the weakest, further highlighting the limitations of transformers without additional optimization. In contrast, the MobileViTv3 family, particularly MobileViTv3-S, shows the advantage of hybrid architectures, offering the best balance between efficiency and accuracy. Even the smaller versions, MobileViTv3-XS and XXS, perform well, making them suitable for resource-constrained situation while still providing competitive performance.

Table II presents the affect of various smoothing techniques on the performance of the proposed method for classifying breast ultrasound images into benign and malignant. Firstly, without applying any smoothing, the model achieved satisfactory results in classifying between the two types of lesions. When the MA smoothing technique was applied, there was a slight improvement in the model's ability to accurately classify the ultrasound images.

However, the most important advancement was observed when the TVA smoothing technique employed. TVA leads to a significant improvement in the classification performance, enhancing the accuracy, precision, recall, and F1-score by approximately 2%. This suggests that TVA allows the model to more effectively capture and utilize the subtle variations in ultrasound imaging data, which is critical for differentiating between benign and malignant lesions. TVA's ability to smooth

TABLE I. RESULTS OF THE PROPOSED METHOD WITH VARIOUS BACKBONE FEATURE EXTRACTORS WITHOUT SMOOTHING

Method	Accuracy	Precision	Recall	F1-score
ResNet-152 [31]	85.91	85.18	84.66	84.91
MobileViTv3-S [21]	89.33	88.90	88.79	88.84
MobileViTv3-XS [21]	86.55	84.61	84.02	84.31
MobileViTv3-XXS [21]	83.76	81.97	80.35	81.15
ConvNeXt V2 [32]	88.10	88.22	86.05	87.12
DEiT [33]	86.75	85.82	84.95	85.38
BEiT v2 [34]	85.46	84.24	83.74	83.98
XCiT [35]	83.67	82.54	81.78	82.15

the data while preserving key features seems to help the model focus on more relevant regions, thus improving overall diagnostic accuracy and reducing the risk of misclassification, which is crucial in breast cancer detection.

TABLE II. RESULTS OF THE PROPOSED METHOD WITH DIFFERENT SMOOTHING METHODS

Method	Accuracy	Precision	Recall	F1-score
W/o smoothing	89.33	88.90	88.79	88.84
MA	89.56	88.94	88.86	88.89
TVA	**91.58**	**90.76**	**90.11**	**90.43**

Table III compares the performance of the proposed method with and without the use of quality weights, evaluating key metrics. The inclusion of quality weights clearly improves performance across all metrics. When quality weights are not applied, the method achieves an accuracy of 91.58%, with a precision of 90.76%, recall of 90.11%, and F1-score of 90.43%. These are strong results, indicating that the model can effectively make predictions, but there is room for improvement in its ability to generalize and balance precision and recall.

TABLE III. RESULTS OF THE PROPOSED METHOD WITH AND WITHOUT QUALITY WEIGHTS

Method	Accuracy	Precision	Recall	F1-score
w/o quality weights	91.58	90.76	90.11	90.43
with quality weights	**93.78**	**93.65**	**92.94**	**93.29**

When quality weights are introduced, the performance improves substantially, with accuracy increasing to 93.78%, a significant boost over the baseline. Similarly, precision rises to 93.65%, recall to 92.94%, and the F1-score to 93.29%. This improvement can be attributed to the model's ability to assign higher importance to more informative lesion or tumor related features during training, resulting in more refined feature extraction and better classification outcomes. The use of quality weights enhances the model's ability to focus on higher-quality data, leading to better overall predictions and higher consistency in its results. Fig. 4 shows the area under the receiver operating characteristic (AUROC) scores of 0.94 and 0.97 for the model without and with quality weights, respectively.

Fig. 5 shows the explainability of the proposed model using GradCam and LIME. The red refers to a higher probability of the presence of lesion or tumor, while the blue represents a lower probability of the existence of background region. Based on visual inspection, the model correctly identified both benign lesions [Fig. 5(a), (b)] by focusing on hypoechoic regions.

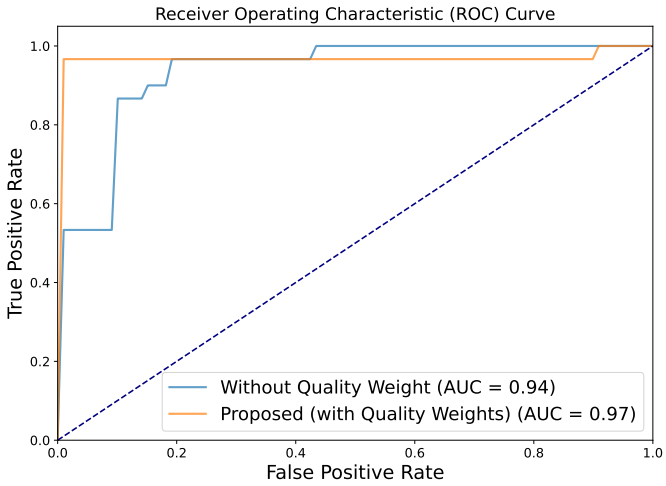


Fig. 4. The AUROC curves of the proposed method with and without quality weights.

Additionally, small tumors [Fig. 5(c), (d)] with ambiguous boundaries were accurately classified as malignant, with the model highlighting critical regions while ignoring background pixels.

C. Comparisons with Related Methods and Discussion

Table IV compares the performance of the proposed method with an existing method from the literature, specifically the method from [10]. The proposed method achieves the highest performance across all metrics, with an accuracy of 93.78%, precision of 93.65%, recall of 92.94%, and F1-score of 93.29%. These results demonstrate a clear improvement over the existing method, which, while still effective, yields slightly lower accuracy (91.66%) and F1-score (92.33%).

The improved performance of the proposed method can be attributed to its ability to better capture important features and balance precision and recall. The higher F1-score indicates that the proposed method handles the trade-off between precision and recall more effectively, resulting in more accurate and reliable predictions. In comparison, while [10] method performs well, it falls short in terms of overall accuracy and F1-score, suggesting that the proposed method offers a more refined approach to the problem.

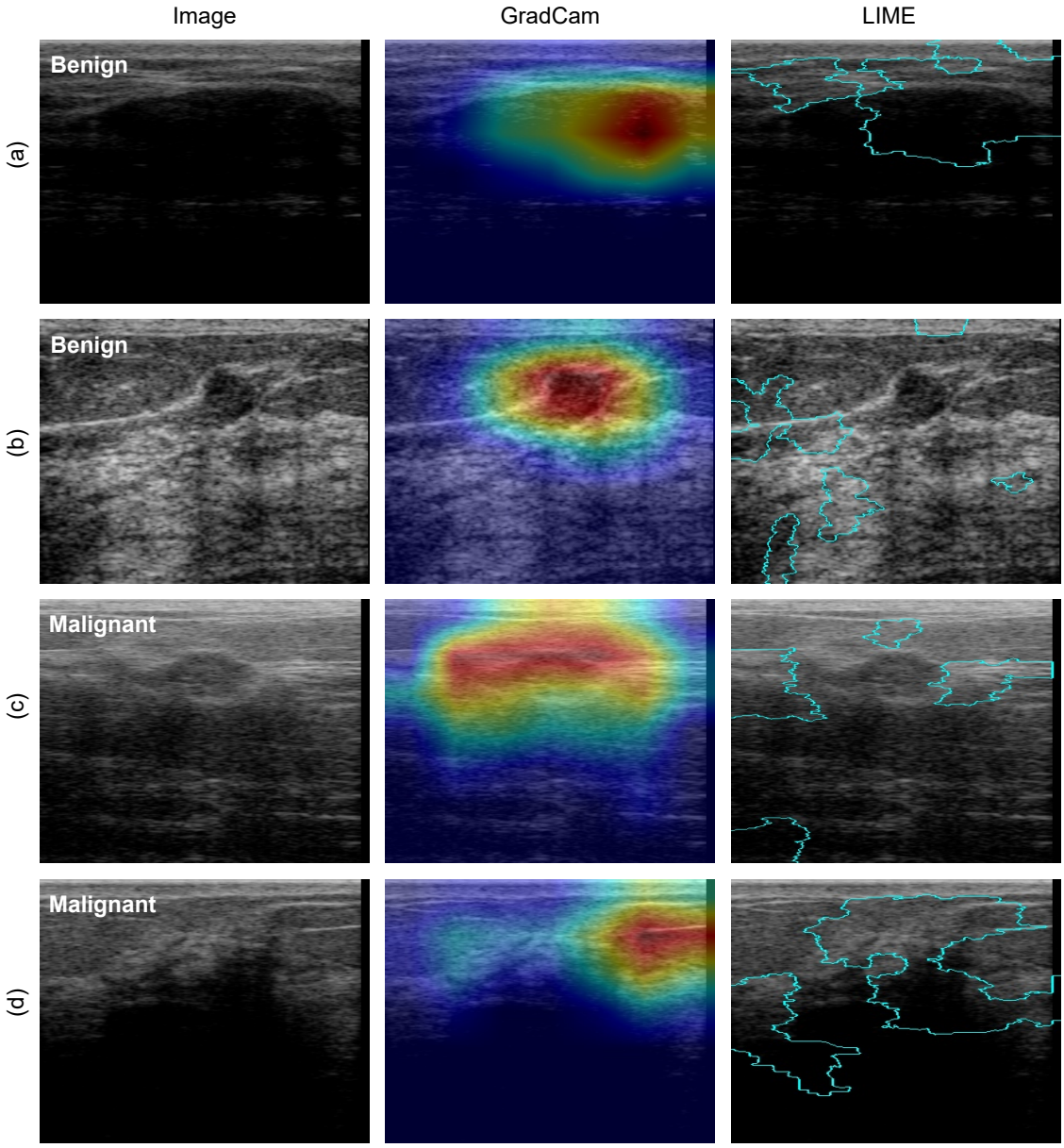


Fig. 5. Explanation of the proposed model using Grad-CAM [29] and LIME [30] methods. (a,b) benign cases, and (c,d) malignant cases.

TABLE IV. STATE-OF-THE-ART RESULTS COMPARISON

Method	Accuracy	Precision	Recall	F1-score
Proposed	**93.78**	**93.65**	**92.94**	**93.29**
[10]	91.66	93.05	92.69	92.33

V. Conclusion and Future Work

This paper presents a novel approach to breast tumor classification using dynamic pooling of BUS sequences, combining the strengths of both CNN and transformer architectures. By incorporating weighted dynamic pooling based on image quality metrics, such as blurriness and brightness, our method effectively mitigates the impact of noise and artifacts commonly found in BUS images. Comprehensive experiments demonstrate that our approach, particularly when using MobileViTv3-S, significantly outperforms existing methods, achieving an accuracy of 93.78%. The inclusion of quality weights further enhances classification performance, highlighting the importance of prioritizing high-quality image frames. Not only does our model achieve higher accuracy, but it also provides better interpretability through Grad-CAM visualizations, facilitating the understanding of tumor characteristics. The results suggest that our approach can offer a robust, reliable, and interpretable solution for breast cancer detection in clinical settings.

One limitation of this study is the reliance on a single BUS video sequence dataset to evaluate the efficacy of the proposed method. Additionally, the small sample size of the dataset presents another limitation.

Future work will focus on integrating additional ultrasound modalities (e.g. BUS and CEUS) with the proposed method to further enhance classification accuracy. Additionally, larger and more diverse datasets will be collected to improve the robustness and performance of the developed classification models.

Acknowledgment

This research was partly supported by Spanish Government through Project TED2021-130081B-C21, Project PDC2022-133383-I00, and Project PID2019-105789RB-I00.

References

[1] J. D. B. Fuentes, E. Morgan, A. de Luna Aguilar, A. Mafra, R. Shah, F. Giusti, J. Vignat, A. Znaor, C. Musetti, C.-H. Yip *et al.*, "Global stage distribution of breast cancer at diagnosis: a systematic review and meta-analysis," *JAMA oncology*, 2024.

[2] R. L. Siegel, A. N. Giaquinto, and A. Jemal, "Cancer statistics, 2024." *CA: a cancer journal for clinicians*, vol. 74, no. 1, 2024.

[3] M. I. Tsarouchi, A. Hoxhaj, and R. M. Mann, "New approaches and recommendations for risk-adapted breast cancer screening," *Journal of Magnetic Resonance Imaging*, vol. 58, no. 4, pp. 987–1010, 2023.

[4] C. Shan, T. Tan, J. Han, and D. Huang, "Ultrasound tissue classification: a review," *Artificial Intelligence Review*, vol. 54, no. 4, pp. 3055–3088, 2021.

[5] K. Malherbe and D. Tafti, "Breast ultrasound," *StatPearls*, 2024.

[6] M. Bahl, J. M. Chang, L. A. Mullen, and W. A. Berg, "Artificial intelligence for breast ultrasound: Ajr expert panel narrative review," *American Journal of Roentgenology*, 2024.

[7] J. Ellis, K. Appiah, E. Amankwaa-Frempong, and S. C. Kwok, "Classification of 2d ultrasound breast cancer images with deep learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 5167–5173.

[8] G. Işık and İ. Paçal, "Few-shot classification of ultrasound breast cancer images using meta-learning algorithms," *Neural Computing and Applications*, pp. 1–13, 2024.

[9] M. G. Lanjewar, K. G. Panchbhai, and L. B. Patle, "Fusion of transfer learning models with lstm for detection of breast cancer using ultrasound images," *Computers in Biology and Medicine*, vol. 169, p. 107914, 2024.

[10] M. A. Hassanien, V. K. Singh, D. Puig, and M. Abdel-Nasser, "Predicting breast tumor malignancy using deep convnext radiomics and quality-based score pooling in ultrasound sequences," *Diagnostics*, vol. 12, no. 5, p. 1053, 2022.

[11] M. Abdel-Nasser, J. Melendez, A. Moreno, O. A. Omer, and D. Puig, "Breast tumor classification in ultrasound images using texture analysis and super-resolution methods," *Engineering Applications of Artificial Intelligence*, vol. 59, pp. 84–92, 2017.

[12] S. Sushanki, A. K. Bhandari, and A. K. Singh, "A review on computational methods for breast cancer detection in ultrasound images using multi-image modalities," *Archives of Computational Methods in Engineering*, vol. 31, no. 3, pp. 1277–1296, 2024.

[13] A. Sahu, P. K. Das, and S. Meher, "An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images," *Biomedical Signal Processing and Control*, vol. 87, p. 105377, 2024.

[14] K. S. Rao, P. V. Terlapu, D. Jayaram, K. K. Raju, G. K. Kumar, R. Pemula, V. Gopalachari, and S. Rakesh, "Intelligent ultrasound imaging for enhanced breast cancer diagnosis: Ensemble transfer learning strategies," *IEEE Access*, 2024.

[15] M. Ragab, A. O. Khadidos, A. M. Alshareef, A. O. Khadidos, M. Altwijri, and N. Alhebaishi, "Optimal deep transfer learning driven computer-aided breast cancer classification using ultrasound images," *Expert Systems*, vol. 41, no. 4, p. e13515, 2024.

[16] C. He, Y. Diao, X. Ma, S. Yu, X. He, G. Mao, X. Wei, Y. Zhang, and Y. Zhao, "A vision transformer network with wavelet-based features for breast ultrasound classification," *Image Analysis and Stereology*, vol. 43, no. 2, pp. 185–194, 2024.

[17] Z. Yang, X. Gong, Y. Guo, and W. Liu, "A temporal sequence dual-branch network for classifying hybrid ultrasound data of breast cancer," *Ieee Access*, vol. 8, pp. 82 688–82 699, 2020.

[18] G. Zhao, D. Kong, X. Xu, S. Hu, Z. Li, and J. Tian, "Deep learning-based classification of breast lesions using dynamic ultrasound video," *European Journal of Radiology*, vol. 165, p. 110885, 2023.

[19] M. Han, D. Guo, J. Yuan, and C. Lu, "A spatio-temporal feature fusion network for intelligent analysis of breast cancer contrast-enhanced ultrasound video," in *Proceedings of the 2024 4th International Conference on Bioinformatics and Intelligent Computing*, 2024, pp. 270–274.

[20] Y. Zhang, D. Kong, J. Li, T. Yang, F. Yao, and G. Yang, "Using segment-level attention to guide breast ultrasound video classification," in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2024, pp. 1–5.

[21] S. N. Wadekar and A. Chaurasia, "Mobilevitv3: Mobile-friendly vision transformer with simple and effective fusion of local, global and input features," *arXiv preprint arXiv:2209.15159*, 2022.

[22] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, pp. 199–222, 2004.

[23] V. K. Singh, B. Kucukgoz, D. C. Murphy, X. Xiong, D. H. Steel, and B. Obara, "Benchmarking automated detection of the retinal external limiting membrane in a 3d spectral domain optical coherence tomography image dataset of full thickness macular holes," *Computers in Biology and Medicine*, vol. 140, p. 105070, 2022.

[24] L. Francis and N. Sreenath, "Pre-processing techniques for detection of blurred images," in *Proceedings of International Conference on Computational Intelligence and Data Engineering*, 2019, pp. 59–66.

[25] S. Bezryadin, P. Bourov, and D. Ilinih, "Brightness calculation in digital image processing," in *International Symposium on Technologies for Digital Photo Fulfillment*, vol. 2007, no. 1, 2007, pp. 10–15.

[26] M. A. Hassanien, V. Kumar Singh, D. Puig, and M. Abdel-Nasser, "Transformer-based radiomics for predicting breast tumor malignancy score in ultrasonography," in *Artificial Intelligence Research and Development*. IOS Press, 2022, pp. 298–307.

[27] B. Gheflati and H. Rivaz, "Vision transformers for classification of breast ultrasound images," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2022, pp. 480–483.

[28] V. K. Singh, E. M. Mohamed, and M. Abdel-Nasser, "Aggregating efficient transformer and cnn networks using learnable fuzzy measure for breast tumor malignancy prediction in ultrasound images," *Neural Computing and Applications*, vol. 36, no. 11, pp. 5889–5905, 2024.

[29] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[30] M. T. Ribeiro, S. Singh, and C. Guestrin, "" why should i trust you?" explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[32] S. Woo, S. Debnath, R. Hu, X. Chen, Z. Liu, I. S. Kweon, and S. Xie, "Convnext v2: Co-designing and scaling convnets with masked autoencoders," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 16 133–16 142.

[33] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," in *International conference on machine learning*. PMLR, 2021, pp. 10 347–10 357.

[34] Z. Peng, L. Dong, H. Bao, Q. Ye, and F. Wei, "Beit v2: Masked image modeling with vector-quantized visual tokenizers," *arXiv preprint arXiv:2208.06366*, 2022.

[35] A. Ali, H. Touvron, M. Caron, P. Bojanowski, M. Douze, A. Joulin, I. Laptev, N. Neverova, G. Synnaeve, J. Verbeek *et al.*, "Xcit: Cross-covariance image transformers," *Advances in neural information processing systems*, vol. 34, pp. 20 014–20 027, 2021.

Classification of Moroccan Legal and Legislative Texts Using Machine Learning Models

Amina BOUHOUCHE[1], Mustapha ESGHIR[2], Mohammed ERRACHID[3]
Faculty of Sciences, Mohammed V University in Rabat, LabMIA-SI Rabat, Morocco[1,2]
Regional Center for Education and Formation Professions (CRMEF), Rabat, Morocco[3]

***Abstract*—Artificial intelligence tools have revolutionized many fields, bringing significant progress in automating tasks and solving complex problems. In this article, we focus on the legal domain, where the data to be processed are specific and in large quantities. Our study consists in carrying out an automatic classification of Moroccan legal and legislative texts in Arabic. In addition, we will conduct a series of experiments to evaluate the impact of stemming, class imbalance and the impact of data quantity on the performance of the models used. Given the specificity of the Arabic language, we used Natural Language Processing (NLP) tools adapted to this language. For classification, we worked with the following models: Support Vector Machine (SVM), Random Forests (RF), K Nearest Neighbors (KNN) and Naive Bayes (NB). The results obtained are very impressive, and the comparison of model outputs enriches the debate on specificities of each model.**

***Keywords*—Classification Arabic text; natural language processing; legal data; machine learning**

I. Introduction

Approaching textual data in the Arabic legal field presents dual challenges due to linguistic and legal complexities. The Arabic language is an extremely rich Semitic language both linguistically and culturally. It is the official language in over 20 countries and serves as one of the six official languages of the United Nations (UN). Textual data in the Arabic language presents a challenge due to the language's specificity and complexities. These complexities are evident in the distinction between Literary Arabic comprising Classical Arabic and Modern Standard Arabic (MSA), and Dialectal Arabic, which is highly diverse due to regional and national variations. Furthermore, the linguistic richness of the Arabic language imparts complexity at the syntactic, morphological, semantic, and orthographic levels [1]. Regarding the legal domain, which is a highly interesting field and happens to be the focus of our study, it has been approached using machine and deep learning techniques in diverse languages to explore prediction of judicial decisions [2], synthesis [3], and translation of legal texts [4]. However, not as much research has been done on the subject of Arabic. NLP tools, machine learning, and deep learning models are applied to legal texts to automate certain tedious tasks and enable professionals to focus on high-value-added tasks. In this paper, we have focused on applying these techniques to classify legal texts related to Moroccan regulations and legislation. Several models have been trained, including SVM, RF, KNN and NB classifiers. The aim of our study is to work with a wide range of data and to evaluate the behavior of the models in relation to different situations. We have chosen Moroccan legal and legislative texts covering different fields. These texts include the Civil Status Law, the Organic Finance Law, the Judicial Organization Law, the Traffic Code, the Family Code, the Criminal Procedure Law and the Commercial Code. Each text deals with distinct subjects and uses terms specific to those subjects. To our knowledge, this corpus has not been approached before.

The approach adopted in this work concerns the evaluation of the impact of stemming and data quantities on model performance. According to our information, previous work has not dealt with these aspects when approaching legal data in Arabic. By comparing the performance of models trained on stemmed and non-stemmed data, we can evaluate the behavior of each model in relation to the application or non-application of the stemming approach. In this study, we also analyzed the impact of class imbalance on model performance. Thus, we first evaluated the results obtained when training the models on a dataset containing minority classes, then we trained the models on a dataset where the classes are more balanced. Furthermore, we examined the effect of varying data quantity on the performance of these classifiers. By systematically varying the size of our dataset, we were able to observe how each model responded to different volumes of training data. This analysis provides valuable information on the scalability and robustness of each classifier.

To carry out this study, we have structured our article as follows: The characteristics and specifications of the Arabic language are discussed in Section II. Section III is devoted to the state of related research, while the Section IV outlines the methodology followed and details the experiments conducted in this study. The results obtained are presented and discussed in the Section V. In the conclusion we summarized the results obtained and proposed new perspectives for future work to tackle new legal tasks and enrich the current state of the art.

II. Arabic Language: Characteristics and Specificities

The Arabic language incorporates several particularities that differentiate it from Western languages. In addition to linguistic complexities, Arabic is distinguished by its writing direction, from right to left, and the use of diacritical marks that determine the phonological meaning of a word [5]. The language also includes additional characters, such as "إ، آ، أ، ء، ئ، ؤ، ى" and "ة", whose basic form are the Arabic letters "ا" and "ه" respectively. Furthermore, the syntax of the Arabic language differs from that of Western languages, especially concerning word order. For example, in the sentence "أصدرت المحكمة الحكم" where the direct translation is "issued

the court the judgment", the order of the subject and the verb is reversed compared to their order in the English language, where it is "the court issued the judgment". A deep understanding of grammatical and syntactical structures is essential for an enhanced mastery of the language and to ensure a smooth translation between multiple languages.

Morphologically, a trilateral root has the ability to generate multiple derivatives by adding prefixes, suffixes, or modifying vowels. For instance, the trilateral root "حكم" meaning "to judge" can give rise to various words related to the field of judgment and law. These include "حُكْم" (hukm) meaning "judgment", "حكمة" (hik-ma) meaning "wisdom" and "محكمة" (mah-ka-ma) meaning "court". Derivatives also provide the opportunity to express various nuances, such as gender, with "حاكم" (ha-kim) meaning male judge and "حاكمة" (ha-ki-ma) meaning female judge, as well as the dual form with "حاكمان" (ha-ki-ma-an) or plurals "حكام" (hu-kka-m), and so forth.

Regarding the semantic aspect, there are words that are written in the same way but differ in terms of meaning and pronunciation. Understanding the correct sense and determining the correct pronunciation are made possible by the context of use. The term "شهادة" for example, can have various meanings such as "testimony" or "certificate," and only the context of the sentence or surrounding discourse allows for determining the specific meaning.

The spelling of Arabic letters is also very specific. The shape of the letters changes according to their position in the word. The letter "ع" for example, has four forms: one form at the beginning of the word as in "عقد", a middle form as in the word "تعلم", and two forms at the end of the word as in the words "استماع" and "بيع."

III. Related Works

The use of machine learning tools has gained momentum in the approach to judicial data. Aleteras et al [2] conducted the first study to combine machine learning and NLP tools to predict judicial decisions regarding the violation of human rights convention articles. This study used the Support Vector Machine (SVM) classifier and served as a reference for several subsequent works, particularly those conducted by Sulea et al. [6] and Liu and Chen [7]. In 2017, Katz et al. [8] presented a study in a generalized, consistent, and out-of-sample framework using the Random Forest classifier to predict US Supreme Court behavior. The promising outcomes in predicting judicial case have piqued the interest of researchers worldwide. For instance, Walt et al. [9] applied the Naive Bayes classifier to German jurisdiction, particularly German tax law cases. Likewise, Virtucio et al. [10] utilized Aletras's method to predict decisions of the Philippine Supreme Court. On the other hand, deep learning models were initially experimented by the Chinese for predicting legal judgments as a multitask problem. Luo et al.'s study [11] focused on simultaneously modeling relevant law extraction and charge prediction. Zhong et al. [12] opted for topological learning for predicting three dependent subtasks: applicable law articles, charges, and penalties. On the other hand, Long et al. [13] developed a model based on legal reading comprehension to express complex semantic interactions between factual descriptions, pleadings, and law articles. Ye et al.'s [14] work revolves around a Seq2Seq model to combine charge prediction and court opinion generation. Meanwhile, Hu et al [15] designed a model focused on predicting infrequent charges.

The deep learning models have attracted the attention of other researchers who have worked on datasets from different courts and jurisdictions. Kowsrihawat et al.[16] utilized the BI-GRU model with an attention mechanism to analyze a corpus of criminal cases from the Thai Supreme Court and predict the guilt or innocence of an accused. Chalkidis et al. [17] also worked on data from the European Court of Human Rights (ECHR) for prediction purposes. They tested several models, including SVM BOW, BIGRU-Attention, Hierarchical Attention Network, Label-Wise Attention Network, and HIERARCHICAL-BERT. Their analysis showed that deep learning models perform better. For their part, Malik et al. [18] designed a prediction system for Indian Supreme Court cases, with the capability of explaining the obtained predictions. For prediction, the authors explored classical, sequential, and hierarchical models as well as transformer-based models.

The application of machine learning methods to judicial data in various languages has led to great interest in the exploration of Arabic court data. Study [19] focused on predicting Arabic judgments from the Errachidia court in Morocco, specifically targeting accident cases. Deployed machine learning models included Linear Regression, Random Forest, and Decision Trees, achieving encouraging results with 91% accuracy for Random Forest. Shamma et al. [20] addressed information extraction from Arabic legal documents, applying Arabic-specific natural language processing techniques and machine learning models like SVM, Decision Trees, and KNN. In study [21], researchers examined the prediction of verdicts, legal articles, and probabilities of judgment, focusing on child custody and marriage annulment cases in the Arabic language. They experimented with TF-IDF and word2vec representation techniques and used various machine and deep learning models including SVM, Logistic Regression (LR), LSTM, and BiLSTM. Article [22] introduces TaSbeeb, an innovative decision support tool for the Saudi judicial system. Designed for Arabic language, it retrieves judicial reasoning and performs multiclass classification of legal cases. The study proposes deep models Att-GRU, BiLSTM, and BiGRU with stacking approach, both homogeneous and heterogeneous, along with an Arabic judicial language model called Jud_RoBERTa.

In general, the research works that have approached the legal field in Arabic remain limited. Table I below presents some of these works. Most of them focus on judgments or judicial decisions in narrow fields. In addition, this type of data can sometimes lack precision, which can have an impact on the final decision given by the models. Another limitation encountered by the authors lies in the quantities of data available, whereas the models require adequate and representative quantities of data for better training. In our study we worked with legal and legislative data covering seven different domains. In addition, we focused on the impact of data on model performance by changing the amounts of data used and evaluating the impact of class imbalance and

stemming on model performance.

TABLE I. STUDIES ON THE APPLICATION OF ARTIFICIAL INTELLIGENCE APPROACHES TO ARABIC LEGAL DATA

Articles	Experiment	Data
[19]	Predicting the outcome of accident cases	Accident cases issued by the Errachidia court in Morocco
[20]	Information extraction from Arabic legal documents	Appeal cases for leased premises lawsuits
[21]	Predicting judgments and legal articles or proof as well as probabilities of expected results	Marriage annulment and child custody cases
[22]	Development of a decision support system for the Saudi judicial court	Family and personal status cases
[25]	Classification of Arabic legal documents	Judgments relating to real estate and road traffic issued by the Moroccan Supreme Court

IV. METHODOLOGY

The categorization of textual data is one of the fundamental pillars of much research that has explored the contributions of artificial intelligence tools to the legal field. This importance stems from the need to optimize time and resources, particularly in view of the large quantities of data to be processed, including judgments, legal texts and investigation reports. The classification of this type of data provides invaluable assistance to legal professionals, facilitating access to relevant information, comparative analysis between different laws or regulations, and research into previous similar cases.

In this study, we focused on Arabic-language legal data relating to Moroccan regulations. To classify this text data, it is important to follow a clearly defined process, including data collection and pre-processing, vector representation of the texts, model application and performance evaluation (Fig. 1).

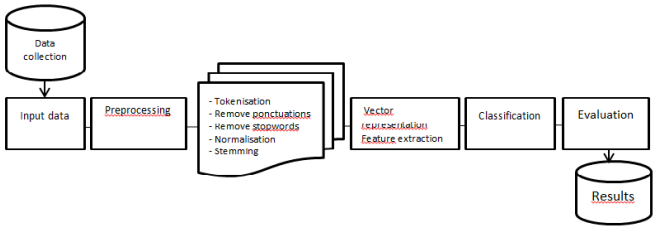


Fig. 1. Classification process for legal and legislative texts.

A. Data Collection

The dataset represents the input for algorithms and the foundation of learning. The models' performance depends entirely on the quality and quantity of the chosen corpus.

In this study, we have gathered seven Moroccan legislative and legal texts, detailed in (Table II). These texts include articles related to the Civil Status Law, the Organic Finance Law, the Judicial Organization Law, the Traffic Code, the Family Code, the Criminal Procedure Law and the Commercial Code. The number of articles varies from one legal text to another, and similarly, the number of words in these articles varies from one article to another. The total number of articles is 2510. The Commercial Code and the Criminal Procedure Law are the most extensive, with 798 and 757 articles, respectively. Since the assignment of articles to a class is determined by the designation of the appropriate legal text, the distribution of classes in our dataset is not balanced.

TABLE II. OVERVIEW OF STUDIED DATA

Legal text	Number of articles
"القانون رقم " 36.21 "المتعلق بالحالة المدنية"	59
"القانون التنظيمي رقم " 130-13 "قانون المالية"	70
" القانون رقم " 38.15 "المتعلق بالتنظيم القضائي"	111
"مدونة السير على الطرق"	318
"مدونة الأسرة"	400
"القانون رقم " 22.01 "المتعلق بالمسطرة الجنائية"	757
"مدونة التجارة"	798

B. Preprocessing

Given our interest in the Arabic legal domain, we have examined NLP tools applied to Arabic-language data. Below, we present the main pre-processing steps required to prepare textual data and make it easy and suitable to handle.

1) Tokenization: Segmentation of text into meaningful linguistic entities.

2) Removing stop words: Elimination of words that don't provide significant information, using the Arabic stop words list.

3) Normalization: Normalization involves writing words in a specific form, bringing letters back to their base form. For example, converting "hamza" in all its forms "إ، آ، أ، ء، ئ، ؤ، ى" to "alef" "ا".

4) Stemming: The process of reducing words to their base form. Among the most frequently employed stemmers for the Arabic language, there are Khoja stemmer and ISRI Stemmer [23]. In this study, we used ISRI Stemmer.

	Sentence1	Sentence 2
Input text	يأمر قاضي التحقيق بتبليغ الشكاية إلى وكيل الملك أو الوكيل العام للملك لتقديم ملتمساته.	تشتمل النفقات المتعلقة بالدين العمومي على النفقات من فوائد وعمولات والنفقات المتعلقة باستهلاكات الدين المتوسط والطويل الأجل
Pre-processed text without stemming step	يامر قاضي التحقيق بتبليغ الشكايه وكيل الملك الوكيل العام للملك لتقديم ملتمساته	تشتمل النفقات المتعلقه بالدين العمومي النفقات فوائد وعمولات والنفقات المتعلقه باستهلاكات الدين المتوسط والطويل الاجل
Pre-processed text with stemming step	امر قضي حقق بلغ شكي وكل ملك وكل عام ملك قدم ملتمساته	شمل نفق تعلق دين عمم نفق فئد عمل نفق تعلق باستهلاك دين توسط طول اجل

Fig. 2. Examples of sentences illustrating the pre-processing of legal and legislative texts.

Fig. 2 illustrates the pre-processing of two sentences, the first extracted from the Criminal Procedure Law and the second from the Organic Finance Law. The results of the pre-processing show that the tools applied have succeeded to a

certain extent in approximating the Arabic text. However, the difficulties encountered include the case of attached words, where it is difficult to identify stop words. We take as an example the word "باستهلاك" composed of two elements "ب" which is a preposition and the word "استهلاك" meaning "consumption". In principle the stop word should be removed. And the term "استهلاك" should undergo stemming. By analyzing the output of the preprocessing with stemming and without stemming step, we noticed that the tool used has succeeded in reducing words to their roots, such as transforming "التحقيق" into "حقق" and "الوكيل" into "وكل" . However the stemming of some words failed. For example the word "فوائد" was rendered to a non-existent word in the Arabic language "فئد" and the words "ملتمساته" and "استهلاك" were not stemmed.

C. Vector Representation Methods

To ensure that textual data is readable and interpretable by algorithms, it must be converted into vector representations. The methods commonly used for this conversion are as follows:

1) BOW: Vector representation of the document that associates each word with its frequency of occurrence.

2) TFIDF: A representation that emphasizes the importance of a word in a document relative to its occurrence in the entire corpus.

$$TF - IDF(term) = TF(term) \times log(\frac{N}{DF(term)}) \quad (1)$$

TF: the number of times the term occurs in a document

DF: the number of documents in which the term is present

N : total number of documents

3) Word embedding: Representations that capture semantic relationships between words. These representations are generated by widely used models such as Word2Vec, GloVe and FastText [24], initially designed for Western languages and applied to the Arabic language.

D. Model Application

Machine learning models are selected according to the nature of the data and the specific objectives of the task. In this study, we chose to use four of the best-known classical classification models: SVM, KNN, Random Forest and Naive Bayes, based on the scikit-learn implementation.

1) Support Vector Machine (SVM): A supervised learning method employed for classification tasks. SVM focuses on choosing the optimal hyperplane to maximize margin separation and incorporates the Kernel function. This function facilitates the projection of data into a higher-dimensional space to handle non-linear classification challenges. The SVM classifier has proven to be very interesting for text classification [25]. Its application to Arabic legal data has been conducted for several tasks, including classification [21] and information retrieval [20].

2) K Nearest Neighbors (KNN): A straightforward and non-parametric classification model based on neighborhood principle. It depends on selecting a distance metric to calculate the distance between instances. The classification of a new instance is decided by the majority class of the k nearest data points in the training set. Although this technique is widely used in text classification, its application to Arabic legal data remains limited, among the few works that have used it, we can cite the study by Ikram and Chakir [25] to classify Moroccan court rulings.

3) Naive bayes: A probabilistic model grounded in Bayes' theorem with conditional independence assumption. This means that the values of predictors of a given class are conditionally independent of each other.

$$P(c_k \mid x_i) = \frac{P(c_k) \times \prod_{j=1}^{n} P(x_{ij} \mid c_k)}{\sum_{s=1}^{K} P(c_s) \prod_{j=1}^{n} P(x_{ij} \mid c_s)} \quad (2)$$

K: number of class modalities
$C = (c_1, c_2, ...c_k, ..., c_K)$: the class variable
n: the number of features or explanatory variables

$x_i = (x_{i1}, ...x_{ik}...x_{in})$: vector of explanatory variables

$P(c_k \mid x_i)$: the probability that the document represented by x_i, belongs to the class c_k.

$P(x_{ij} \mid c_k)$: The likelihood of the feature x_{ij} conditioned on the class c_k

P(c_k): represents the prior probability associated with the class c_k

The target class is the class for which the probability P($c_k \mid x_i$) is maximum:

$$c_k = ArgMax(P(c_k) \prod_{j=1}^{n} P(x_{ij} \mid c_k) \quad (3)$$

The Naive Bayes classifier has been widely used in various research exploring the legal domain across multiple languages and it is also employed in the classification of legal texts in Arabic [25][26].

4) Random forest: An ensemble machine learning model that combines decision trees of CART-type through aggregation. Each tree is built from a bootstrap sample of the training set, using a random subset of features. The model's prediction is established through a majority vote from all the trees. Random Forest is a widely used model for text data analysis. It has been employed by researchers such as Katz [8], Liu and Chen [7] and others to approach judicial data. Concerning the Arabic language, this model has been applied to tasks such as isolated Arabic character recognition [27] and legal document classification [25].

E. Evaluation

The performance of classification and prediction models is evaluated through various metrics. This approach aims to analyze the quality of results, the overall efficiency of models, and to understand their behavior in a variety of situations. The most commonly used measurements in this context include:

Accuracy: Proportion of correctly classified cases to the total number of cases.

$$Accuracy = \frac{TP + TN}{Total} \tag{4}$$

Recall: Proportion of cases correctly assigned to a class i to all cases in that class.

$$Rappel = \frac{TP}{TP + FN} \tag{5}$$

Precision: Proportion of cases correctly assigned to class i compared to all cases predicted to be assigned to class i.

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

F1 score: Harmonic mean of precision and recall.

TP: True positive

TN: True negative

FP: False positive

FN: False negative

These performance measures are typically used for binary classification, however, there are methods to generalize them to include multiple classification:

1) Micro approach: The global metrics are calculated from the set contributions of all classes. The TP, TN, FP and FN of all classes are combined to calculate a single accuracy, precision, recall and F1-score.

2) Macro approach: consists of calculating the metrics for each class individually and then calculating an average of all the classes for each measure. These averages can be arithmetic when the classes are balanced and weighted averages when the classes are unbalanced.

F. Experimentation

Due to our interest in applying machine learning tools in the legal field, we conducted a series of experiments to evaluate the impact of stemming, class imbalance and data quantity on model performance.

Our data corpus consists of articles from seven legal and legislative texts (Table II). The designation of the appropriate text constitutes the class assignment for the articles, resulting in a total of seven classes. To ensure the quality of our analysis, we first prepared our data using preprocessing techniques mentioned above. We then applied our models and evaluate the performance obtained. We conducted five distinct experiments to assess the performance of our models in different scenarios. First, we trained classification models on the entire corpus and evaluated performance (Table III). This first experiment served as a basis of comparison for some of the subsequent experiments. In the second experiment, we drew inspiration from the study by Taghva et al. [23]to evaluate the effectiveness of stemming on the performance of our models. In this experiment, we eliminated the stemming step and compared the results (Table IV) with those of the first experiment. Inspired by Sulea et al's study [6] of French-language court decisions, we came up with the idea of assessing the impact of class imbalance and the presence of minority classes when exploring Arabic-language legal documents. Thus, in a third experiment, we assessed the impact of minority classes on model performance by eliminating classes with fewer than 200 articles and retraining our models (Table V). This reduced the corpus to four classes, and the results were compared with those from the first experiment. In the fourth experiment, we evaluated the impact of class imbalance on model performance (Table VI). To do this, we worked with the four classes from the previous experiment and retained 300 articles per class. Finally, to assess the impact of data quantity, we conducted a fifth experiment where we reduced the number of articles to 50 per class(Table VII).

V. Discussion

Our study focuses on the analysis of Moroccan legislative and legal texts using natural language processing tools specifically designed for the Arabic language. We conducted classification experiments on this corpus of data using four classifiers: SVM, KNN, Random Forest and Naive Bayes.

TABLE III. Model Performance with Stemming Step

Classifier	Accuracy	Precision	Recall	F1 score
SVM	94.58	94.9	94.58	94.45
KNN	92.17	92.56	92.17	92.26
Random Forest	91.16	91.91	91.16	90.75
Naive Bayes	92.17	92.72	92.17	91.82

TABLE IV. Model Performance without Stemming Step

Classifier	Accuracy	Precision	Recall	F1 score
SVM	92.57	93.16	92.57	91.86
KNN	92.57	93.16	92.57	92.77
Random Forest	91.97	92.4	91.97	91.75
Naive Bayes	95.38	95.6	95.38	95.19

TABLE V. Model Performance Without Minority Classes

Classifier	Accuracy	Precision	Recall	F1 score
SVM	95.82	95.89	95.82	95.83
KNN	96.26	96.28	96.26	96.26
Random Forest	94.95	94.98	94.95	94.95
Naive Bayes	93.41	93.48	93.41	93.41

TABLE VI. Model Performance on Balanced Classes (300 Articles Per Class)

Classifier	Accuracy	Precision	Recall	F1 score
SVM	96.25	96.38	96.25	96.25
KNN	95.42	95.51	95.42	95.42
Random Forest	93.33	93.54	93.33	93.34
Naive Bayes	95.83	96.04	95.83	95.88

TABLE VII. MODEL PERFORMANCE ON SMALL BALANCED DATA (50 ARTICLES PER CLASS)

Classifier	Accuracy	Precision	Recall	F1 score
SVM	92.5	95.00	92.5	93.01
KNN	95	95.23	95	95.02
Random Forest	92.5	93.75	92.5	92.72
Naive Bayes	97.50	97.86	97.50	97.55

The graph in Fig. 3 summarizes the accuracy of all the models across the five experiments detailed above.

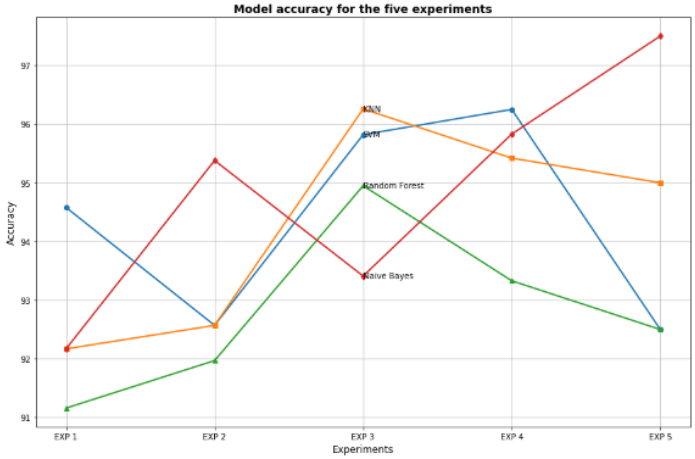


Fig. 3. Model accuracy for the five experiments.

The results of the first experiment (Table III), carried out on the whole corpus, are very promising, exceeding 90% for all models. The SVM classifier achieved the best performance across all metrics, with an accuracy of 94.58% and an F1 score of 94.45%. This finding is consistent with that of Liu and Chen [7], who worked on English-language data from the European Court of Human Rights and confirmed that SVM was the most efficient classifier.

In the second experiment, where we omitted the stemming step (Table IV), we observed that the classifiers behaved differently. The performance of the SVM model decreased, achieving 92.57% accuracy, while the performance of KNN, Random Forest and Naive Bayes has improved, achieving 92.57%, 91.97% and 95.38% accuracy respectively. This difference in behavior is explained by the fact that each classification algorithm has its own characteristics and assumptions. This experiment showed that the KNN, Random Forest and Naive Bayes algorithms are sensitive to the changes introduced by stemming, and are less tolerant of the loss of information it generates. These models are better suited to exploiting the lexical diversity of non-stemmed data, giving them a better understanding of nuances and relationships between words. On the other hand, the SVM classifier, whose performance decreased when trained on non-stemmed words, has been shown to be sensitive to the noise introduced by morphological variations. In addition, the use of stemmed words, allowed the SVM to benefit from the dimensionality reduction introduced by stemming.

The third experiment, in which the minority classes were removed (Table V), revealed an improvement in the performance of all models, with an accuracy of 95,82% for SVM, 96,26% for KNN, 94,95% for Random Forest and 93,41% for Naive Bayes. This suggests that the presence of minority classes was limiting the models' ability to train effectively and identify underlying patterns in the data. By eliminating these minority classes, the models were able to focus on more representative classes, reducing noise and imbalances in the training data. This simplification of the dataset enabled the algorithms to generalize better and produce more accurate and consistent results.

The fourth experiment assessed the impact of class balance, retaining 300 articles per class for the four classes selected in the previous experiment. This approach ensured a fair distribution of data between classes, reducing the bias introduced by class imbalances. By balancing the number of articles per class, the models were able to train on representative datasets, achieving interesting levels of accuracy: 96.25% for SVM, 95.42% for KNN, 93.33% for Random Forest and 95.83% for Naive Bayes (Table VI).
These last two experiments demonstrate the importance of class balance in the machine learning model training process, contributing to more robust performance.

By reducing the amount of data in the fifth experiment (Table VII), we observed a decrease in the performance of the SVM, KNN and Random Forest classifiers compared with the fourth experiment. The accuracy of these classifiers is reduced to 92.5%, 95% and 92.5% respectively. This decrease can be attributed to various factors inherent to these algorithms, mainly the careful tuning of hyperparameters. This tuning process is often made more difficult when the available data is limited. In addition, the Random Forest model, for example, requires a sufficient amount of data for each tree to learn meaningful and diverse patterns. By contrast, the Naive Bayes classifier performed better with small amounts of data. This performance is attributed to the simplicity of the model and its conditional independence assumption, which assumes that each feature contributes independently to the probability of each class.

VI. CONCLUSION

In this study, we were interested in the application of natural language processing and machine learning tools to classify Moroccan legal and legislative data. Given that our datasets are in Arabic, this adds a second level of complexity to our study. The texts under study include the Law on Civil Status, the Organic Law on Finance, the Law on Judicial Organization, the Traffic Code, the Family Code, the Law on Criminal Procedure and the Commercial Code. To the best of our knowledge, this data has not been covered before. Our task was to train the SVM, KNN, Random Forest and Naive Bayes models to assign each article to the appropriate class. In addition to this classification task, we evaluated the impact of stemming, class imbalance and data quantity on the behavior of the models used. In general, our experiments have shown that models perform better with balanced classes, but react differently to stemming and data quantity. Although these classical models generally perform well, it's important to note that they remain sensitive to word variations and the size of the training datasets. In future work, we will explore new perspectives by investigating other models, including deep

learning and language models, in order to tackle new legal tasks.

ACKNOWLEDGMENT

This work was supported by the Ministry of Higher Education, Scientific Research and Innovation, the Digital Development Agency (DDA) and the CNRST of Morocco (Alkhawarizmi/2020/25).

REFERENCES

[1] Shaalan, K., Siddiqui, S., Alkhatib, M., & Abdel Monem, A. (2019). Challenges in Arabic natural language processing. In Computational linguistics, speech and image processing for arabic language (pp. 59-83).

[2] Aletras, N., Tsarapatsanis, D., Preoţiuc-Pietro, D., & Lampos, V. (2016). Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. PeerJ computer science, 2, e93.

[3] Anand, D., & Wagh, R. (2022). Effective deep learning approaches for summarization of legal texts. Journal of King Saud University-Computer and Information Sciences, 34(5), 2141-2150.

[4] Wiesmann, E. (2019). Machine translation in the field of law: A study of the translation of Italian legal texts into German. Comparative Legilinguistics, 37(1), 117-153.

[5] Darwish, K., Habash, N., Abbas, M., Al-Khalifa, H., Al-Natsheh, H. T., Bouamor, H., ... & Mubarak, H. (2021). A panoramic survey of natural language processing in the Arab world. Communications of the ACM, 64(4), 72-81.

[6] Şulea, O. M., Zampieri, M., Vela, M., & van Genabith, J. (2017, September). Predicting the Law Area and Decisions of French Supreme Court Cases. In Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017 (pp. 716-722).

[7] Liu, Z., & Chen, H. (2017, November). A predictive performance comparison of machine learning models for judicial cases. In 2017 IEEE Symposium series on computational intelligence (SSCI) (pp. 1-6). IEEE.

[8] Katz, D. M., Bommarito, M. J., & Blackman, J. (2017). A general approach for predicting the behavior of the Supreme Court of the United States. PloS one, 12(4), e0174698.

[9] Waltl, B., Bonczek, G., Scepankova, E., Landthaler, J., & Matthes, F. (2017). Predicting the outcome of appeal decisions in Germany's tax law. In Electronic Participation: 9th IFIP WG 8.5 International Conference, ePart 2017, St. Petersburg, Russia, September 4-7, 2017, Proceedings 9 (pp. 89-99). Springer International Publishing.

[10] Virtucio, M. B. L., Aborot, J. A., Abonita, J. K. C., Avinante, R. S., Copino, R. J. B., Neverida, M. P., ... & Tan, G. B. A. (2018, July). Predicting decisions of the philippine supreme court using natural language processing and machine learning. In 2018 IEEE 42nd annual computer software and applications conference (COMPSAC) (Vol. 2, pp. 130-135). IEEE.

[11] Luo, B., Feng, Y., Xu, J., Zhang, X., & Zhao, D. (2017, September). Learning to Predict Charges for Criminal Cases with Legal Basis. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (pp. 2727-2736).

[12] Zhong, H., Guo, Z., Tu, C., Xiao, C., Liu, Z., & Sun, M. (2018). Legal judgment prediction via topological learning. In Proceedings of the 2018conference on empirical methods in natural language processing (pp. 3540-3549).

[13] Long, S., Tu, C., Liu, Z., & Sun, M. (2019). Automatic judgment prediction via legal reading comprehension. In Chinese Computational Linguistics: 18th China National Conference, CCL 2019, Kunming, China, October 18–20, 2019, Proceedings 18 (pp. 558-572). Springer International Publishing.

[14] Ye, H., Jiang, X., Luo, Z., & Chao, W. (2018, June). Interpretable Charge Predictions for Criminal Cases: Learning to Generate Court Views from Fact Descriptions. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers) (pp. 1854-1864).

[15] Hu, Z., Li, X., Tu, C., Liu, Z., & Sun, M. (2018, August). Few-shot charge prediction with discriminative legal attributes. In Proceedings of the 27th international conference on computational linguistics (pp. 487-498).

[16] Kowsrihawat, K., Vateekul, P., & Boonkwan, P. (2018, October). Predicting judicial decisions of criminal cases from Thai Supreme Court using bi-directional GRU with attention mechanism. In 2018 5th Asian Conference on Defense Technology (ACDT) (pp. 50-55). IEEE.

[17] Chalkidis, I., Androutsopoulos, I., & Aletras, N. (2019, July). Neural Legal Judgment Prediction in English. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 4317-4323).

[18] Malik, V., Sanjay, R., Nigam, S. K., Ghosh, K., Guha, S. K., Bhattacharya, A., & Modi, A. (2021, August). ILDC for CJPE: Indian Legal Documents Corpus for Court Judgment Prediction and Explanation. arXiv preprint arXiv:2105.13562

[19] Haidar, A., Ahajjam, T., Zeroual, I., & Farhaoui, Y. (2022). Application of machine learning algorithms for predicting outcomes of accident cases in Moroccan courts. Indonesian Journal of Electrical Engineering and Computer Science, 26(2), 1103-1108.

[20] Shamma, S. A., Ayasa, A., & Yahya, A. (2020, October). Information extraction from arabic law documents. In 2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT) (pp. 1-6). IEEE.

[21] Abbara, S., Hafez, M., Kazzaz, A., Alhothali, A., & Alsolami, A. (2023). ALJP: An Arabic Legal Judgment Prediction in Personal Status Cases Using Machine Learning Models. arXiv preprint arXiv:2309.00238.

[22] Almuzaini, H. A., & Azmi, A. M. (2023). TaSbeeb: A judicial decision support system based on deep learning framework. Journal of King Saud University-Computer and Information Sciences, 35(8), 101695.

[23] Taghva, K., Elkhoury, R., & Coombs, J. (2005, April). Arabic stemming without a root dictionary. In International Conference on Information Technology: Coding and Computing (ITCC'05)-Volume II (Vol. 1, pp. 152-157). IEEE.

[24] Dharma, E. M., Gaol, F. L., Warnars, H. L. H. S., & Soewito, B. E. N. F. A. N. O. (2022). The accuracy comparison among word2vec, glove, and fasttext towards convolution neural network (cnn) text classification. J Theor Appl Inf Technol, 100(2), 31.

[25] Ikram, A. Y., & Chakir, L. O. Q. M. A. N. (2019, October). Arabic text classification in the legal domain. In 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS) (pp. 1-6). IEEE.

[26] Jasim, K., Sadiq, A. T., & Abdullah, H. S. (2019, December). A framework for detection and identification the components of arguments in Arabic legal texts. In 2019 First International Conference of Computer and Applied Sciences (CAS) (pp. 67-72). IEEE.

[27] Rashad, M., & Semary, N. A. (2014). Isolated printed Arabic character recognition using KNN and random forest tree classifiers. In Advanced Machine Learning Technologies and Applications: Second International Conference, AMLTA 2014, Cairo, Egypt, November 28-30, 2014. Proceedings 2 (pp. 11-17). Springer International Publishing.

ERCO-Net: Enhancing Image Dehazing for Optimized Detail Retention

Muhammad Ayub Sabir[1], Fatima Ashraf[2]*, Ahthasham Sajid[3],
Nisreen Innab[4], Reem Alrowili[5], Yazeed Yasin[6]
Department of Information Technology, Beijing University of Technology, China[1,2]
Department of Cyber Security-Riphah Institute of Systems Engineering, Riphah International University, Pakistan[3]
Department of Computer Science and Information Systems-College of Applied Sciences, AlMaarefa University, Diriyah, Saudi Arabia[4,5]
Department of Computer Science and Software Engineering, Al Ain University, Al Ain, United Arab Emirates[6]

***Abstract*—Image dehazing is a crucial preprocessing step in computer vision for enhancing image quality and enabling many downstream applications. However, existing methods often do not accurately restore hazy images while maintaining computational efficiency. To overcome this challenge, we propose ERCO-Net a new fusion framework that combines edge restriction and contextual optimization methods. By using boundary constraints, ERCO-Net extend the boundaries that help in protecting the edges and structures of an image. Contextual optimization impacts the final quality of the dehazed image by enhancing smoothness and coherence. We compare ERCO-Net with conventional approaches such as dark channel prior (DCP), All-in-one dehazing network (AoD), and Feature fusion attention network (FFA-Net). The comparative evaluation highlights the effectiveness of the proposed fusion method, providing significant improvement in image clarity, contrast, and colors. The combination of edge restriction and contextual optimization not only enhances the quality of dehazing but also decreases computational complexity, presenting a promising avenue for advancing image restoration techniques. The source code is available at https://github.com/FatimaAyub12/Image-Dehazing-.**

***Keywords*—Image dehazing; edge restriction; contextual optimization; transmission map estimation; haze removal**

I. Introduction

Enhancing visibility in foggy or hazy conditions is a crucial objective in image processing, with picture dehazing playing a key role in achieving this improvement [1], [2]. Dehazing techniques improve image clarity and expose latent details by lowering the effect of ambient scattering. These methods quantify and eliminate haze-generated degradation using mathematical models while analysing pixel intensities and colours [3]. Popular techniques for enhancing image contrast and visibility include the dark channel prior [4] and atmospheric light estimation [5]. Where effective analysis and decision-making depend on clear, detailed images, dehazing finds uses in many disciplines including computer vision, surveillance and remote sensing. [6]. An illustration of a hazy image and its corresponding dehazed version is presented in the Fig. 1.

Early methods for haze reduction mostly depended on either several views of the same image or extra depth information. Notable contributions in this field are shown by several papers, including [2], [1], [3], [4]. Particles in the atmosphere partially polarise light as Schechner et al. [7] noted. The researchers used this finding to propose a method for quickly reducing haze using polarisation to take two pictures from different angles.

Fig. 1. Example of the image with haze (left) and dehazed image (right).

However, despite the advancements in this field, existing dehazing methods often struggle with maintaining a balance between image restoration quality and computational efficiency. Many of these techniques tend to either oversimplify the haze model or fail to preserve critical image details, such as edges and textures, resulting in suboptimal dehazed images with artifacts and color distortions. Furthermore, the increasing complexity of deep learning-based approaches introduces significant computational overhead, making real-time dehazing a challenging task Single image dehazing poses a challenge due to its inherent lack of complete information, making it an under-constrained problem. To address this, the conventional approach involves integrating additional priors or constraints. This paper delves into this notion by establishing an inherent boundary constraint concerning scene transmission. For the purpose of obtaining the illusive transmission parameters, an optimisation framework is constructed using this restriction alongside with a weighted L_1 norm-based contextual optimisation among neighbouring pixels. ERCO-Net operates on a set of fundamental assumptions and demonstrates the capability to produce high-quality, haze-free images with accurate color representation and intricate edge details. Summarizing, the main contributions of this paper are:

*Corresponding authors.

- Firstly, we introduce a novel constraint concerning scene transmission. This straightforward constraint, with its clear geometric interpretation, proves remarkably effective in image dehazing.
- Secondly, we propose a contextual optimization method that allows us to integrate a filter bank into the image dehazing process. These filters play a crucial role in reducing image noise and enhancing significant image features, such as abrupt edges and corners.
- Finally, we present an efficient optimization scheme, enabling rapid dehazing of large-sized images.

The remainder of this paper is organized as: Section II presents a detailed literature review, highlighting existing approaches to image dehazing. Section III introduces the proposed ERCO-Net framework, including edge restriction and contextual optimization techniques. Section IV discusses the methodology behind the edge restriction from the radiance cube and contextual regularization. Section V provides a comprehensive analysis of experimental results and a comparison with other state-of-the-art methods. Finally, Section VI concludes the paper with a summary of findings and future research directions.

II. Literature Review

The related work critically examines seminal works in image processing, atmospheric scattering models, and computational photography. It highlights the strengths and limitations of existing algorithms, emphasizing the need for novel approaches to address challenges in dehazing, such as preserving image details, handling varying haze densities, and improving computational efficiency. By synthesizing insights from diverse scholarly contributions, this review sets the foundation for proposing innovative solutions to enhance visual clarity and fidelity in hazy environments.

A. Network-Based Approaches for Image Dehazing

Convolutional Neural Networks (CNNs) play a pivotal role in modern deep learning-driven dehazing networks. These networks utilize different modules including standard convolution, dilated convolution, multi-scale fusion, feature pyramid, cross-layer connections, and attention mechanisms. Typically, these modules are combined into multiple basic blocks within a dehazing network architecture. This approach aids in comprehending the underlying principles of different dehazing algorithms. To facilitate understanding, the commonly used basic blocks in network architectures are summarized as follows.

1) Standard convolution neural network: Research has demonstrated the effectiveness of employing standard convolution in a sequential manner for constructing neural networks. Hence, it is a common practice to integrate standard convolution into dehazing models alongside other blocks [8], [9], [10], [11].

2) Dilated convolution: In Dilated Fusion, a method is used that enlarges the receptive field without changing the dimensions of the convolution kernel. Various studies [12], [13], [14], [15], [16] have shown its effectiveness in improving global feature extraction. Additionally, integrating convolution layers with varying dilation rates allows for the extraction of features from different receptive fields.

3) Multi-scale fusion: Research has shown that employing Convolutional Neural Networks (CNNs) with multi-scale convolution kernels can effectively extract features across various visual tasks. This approach utilizes convolution kernels of different scales and combines the extracted features, proving beneficial in tasks such as dehazing [17], [18], [19], [20]. Through fusion strategies, these methods achieve multi-scale details essential for image restoration. Typically, during feature fusion, output features obtained from convolution kernels of varying sizes are spatially concatenated or added together.

4) Feature pyramid: In the domain of digital image processing research, the concept of a feature pyramid emerges as a powerful tool. This pyramid allows for the extraction of information at various resolutions from an image. Within the domain of deep learning-based dehazing networks, researchers [21], [22], [23], [24], [25], [11], [26] have employed this strategy within the intermediate of the network layers. Here, the aim is to capture diverse scales of spatial and channel information, enhancing the ability of network to effectively dehaze images. The author in [27] introduces a novel transformer-based architecture for image dehazing that embeds transmission-aware information into the position encoder, addressing the challenge of varying haze densities across different spatial regions. It offers a new method of global and local feature integration for enhanced image clarity.

5) Cross-layer interconnection: To improve information exchange across different layers and strengthen the network's ability to extract features, CNNs frequently use cross-layer connections. In the domain of dehazing networks, three primary types of cross-layer connections emerge: The authors initially proposed ResNet, this approach is widely utilized in [28], [29], [30], [31], [32]. The second type, Dense Connection is applied in the works of [33], [34], [35]. While the third, Skip Connection have been integrated into various architectures [26], [36], [17], [13].

6) Attention dehazing: The attention mechanism has proven to be highly effective in natural language processing research. In the realm of computer vision, two commonly used attention mechanisms are channel attention and spatial attention. Channel attention plays a crucial role in the feature extraction and reconstruction of 2D images. It emphasizes the significant channels in a feature map, allowing the model to concentrate on essential feature details. This targeted approach boosts the efficiency of feature extraction. Meanwhile, spatial attention centers on the variations in feature positions within the map. For example, it can identify and prioritize regions with distinct characteristics, such as areas heavily affected by haze across the entire map. Integrating these attention mechanisms into neural networks has led to significant advancements in various tasks, including image dehazing. Several state-of-the-art dehazing methods [28], [37], [13], [30], [38], [39], [40], [41], [14], [23], [42], [11] have been applied by a number of advanced dehazing methods, which has greatly led to their commendable performance gains. As shown in [43], adversarial auto-augmentation techniques can improve the generalization of dehazing networks. This method adjusts haze density and distribution, allowing the network to generalize better across synthetic and real-world datasets.

The use of deep networks for image dehazing faced some challenges that include increased computational complexity

TABLE I. COMPARATIVE ANALYSIS OF IMAGE DEHAZING TECHNIQUES

Method	Key Features/Approach	Advantages	Limitations	Performance (PSNR / SSIM)
Dark Channel Prior (DCP) [44]	Atmospheric scattering model, dark channel prior used to estimate transmission map	Simple and effective for natural outdoor scenes	Fails with bright sky regions, introduces halo artifacts	16.62 / 0.817 (NHHaze)
All-in-One Dehazing Network (AoD) [8]	CNN-based approach with end-to-end learning	High computational efficiency, suitable for real-time applications	May oversimplify dehazing, causing loss of details	19.06 / 0.85 (NHHaze)
Feature Fusion Attention Network (FFA-Net) [40]	Multi-scale CNN with attention mechanisms for feature fusion	Preserves fine details, improves image contrast	High computational cost, may overfit on synthetic datasets	34.59 / 0.975 (NHHaze)
ERCO-Net (Proposed)	Edge restriction and contextual optimization, boundary constraints on transmission map	Maintains edge details, smooths transitions, reduces computational complexity	Slightly more complex than basic CNN methods	37.61 / 0.991 (NHHaze)

and possible overfitting due to high parameters' count. Moreover, sometimes deep networks do not work well enough for capturing complex atmospheric scattering effects appropriately in order to gain unsatisfactory quality in dehazed images. In contrast, employing edge restrictions and contextual optimization offers distinct advantages. These methods provide explicit guidance on preserving important image features and structural details during dehazing, ensuring more accurate and visually pleasing results. Moreover, they facilitate better incorporation of prior knowledge about haze distribution and scene characteristics, leading to improved generalization and robustness in various environmental conditions. Table I provides a comparative analysis of the key methods discussed in the literature, highlighting their features, advantages, limitations, and performance. This analysis demonstrates the efficiency of the proposed ERCO-Net method compared to existing approaches

III. EDGE RESTRICTION AND CONTEXTUAL OPTIMIZATION

Edge restriction in image processing is closely related to using spatial information, where certain attention is paid to edges and contours of images as needed [45], [46]. It also emphasizes the importance of providing margins and other small details in various image enhancements. More specifically, in the dehazing model, the boundary constraints must be applied to enhance the accuracy of estimates of the transmission map. This map is particularly useful when wanting to determine the extent of the haze that may be obscuring parts of the image [47], [48].

The core concept of the method entails controlling the dehazing operation based on understanding the type of change between objects or parts in the scene. It is mostly represented in equation form as a regularizer added to the energy function within the dehazing algorithm. Due to the priority given to edges and boundaries, the given algorithm risks receiving an infested image that lacks most of its necessary details while at the same time maintaining the clean image as natural as possible.

A. Working of Edge Restriction

The downsizing of the edges in the ERCO-Net is a significant component of the enhanced procedure. Because the transmission map would be required to estimate the edges of the objects, enhance edge preservation by the dehazing algorithm would be necessary [1]. Through the assignment of boundary constraints, the algorithm takes advantage of spatial information to improve the accuracy of the regression equations for the estimation of transmission map. Mathematically, this can be expressed in the following equation:

$$E_{\text{boundary}} = \lambda_b \int \|\nabla I - \nabla J\|^2 \, dx \tag{1}$$

In this context, E_{boundary} denotes the energy associated with boundary constraints. The variable I represents the hazy input image, J stands for the resulting dehazed image, and λ_b is a parameter that weights the boundary constraints appropriately.

B. Contextual Optimization in Dehazing

Contextual Optimization is the another prominent feature of the given approach which is crucial to be considered. It should resolve the issue of enhancing the dehazing process using contextual data at similar effectiveness. Contextual Optimization tends to enhance the neighboring pixels' coherency in an image thus enhancing the quality of the dehazed image [4]. Mathematically, it can be formulated as:

$$E_{\text{context}} = \lambda_c \int \|\nabla^2 J\|^2 \, dx \tag{2}$$

In this context, E_{context} denotes the energy associated with contextual optimization. The variable J represents the dehazed image, and λ_c is a parameter that optimizes this process.

By integrating edge restriction and contextual optimization, ERCO-Net aims to advance single-image dehazing techniques, providing a more robust solution to the challenges posed by atmospheric haze.

The flow of ERCO-Net is illustrated in Fig. 2.

ERCO-Net brings three key improvements. Firstly, introduce a new rule for how light travels through a scene. This rule is straightforward and easy to understand, and it turns out to be remarkably effective in clearing up hazy images. Secondly, introduce a new way to smooth out images by using a set of filters. These filters not only reduce unwanted noise in the images but also make certain features, like sudden changes and corners, stand out more. Lastly, a faster method has been developed to clean up large images, allowing for

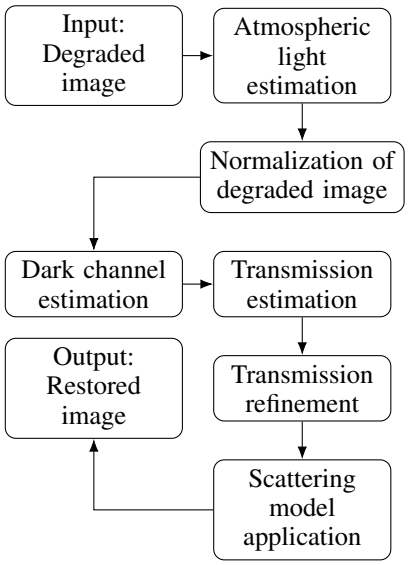


Fig. 2. Flowchart of the ERCO-Net dehazing process.

quick improvements in image clarity without taking much time.

The haze in the image is represented formally in a number of noteworthy work reported in [44], [49], [50], [51], depicted as follows:

$$Img(x) = st(x)R(x) + (1 - st(x))At \tag{3}$$

where, $Img(x)$ is the observed image, $R(x)$ is the scene radiance, At is the global atmospheric light, and $st(x)$ is the scene transmission. It is worthy to mention that the transmission function $st(x) : (0 \leq st(x) \leq 1)$ is related to the image depth.

The key objective of the image dehazing is to recover the image radiance $R(x)$ from $Img(x)$ based on Eq. 3. This requires us to estimate the transmission function $st(x)$ and the global atmospheric light At. Once $st(x)$ and At are estimated, the scene radiance can be recovered by:

$$R(x) = \frac{Img(x) - At}{[\max(st(x), \epsilon)]^{\lambda}} + At \tag{4}$$

IV. Edge Restriction from Radiance Cube

The concept of Boundary Constraint from Radiance Cube is a critical aspect in the context of image dehazing. The Radiance Cube represents the radiance values at different combinations of scene radiance and atmospheric light. The Boundary Constraint is employed to ensure that the estimated radiance values are constrained within physically meaningful bounds. In the context of image dehazing, this helps in preventing unrealistic or exaggerated scene radiance estimations.

- Atmospheric Light Constraint: The atmospheric light (A_t) is constrained to be within the range of 0 to 1. $0 \leq A_t \leq 1$
- Scene Radiance Constraint:
  The estimated scene radiance ($R(x)$) should be non-negative. $R(x) \geq 0$
- Transmission Constraint:
  The transmission ($st(x)$) lies between 0 and 1. $0 \leq st(x) \leq 1$
  4. Radiance Cube Constraint:
  The radiance cube values should be within a valid radiometric range.
  $L_{\text{min}} \leq L(x, \lambda) \leq L_{\text{max}}$
  Here, L_{min} and L_{max} represent the minimum and maximum radiance values, respectively.
- Boundary Constraint Equation:
  Combining the above constraints into a comprehensive boundary constraint equation:

$$R(x) = \frac{Img(x) - A_t}{[\max(st(x), \epsilon)]^{\lambda} + A_t} \tag{5}$$

This equation ensures that the estimated scene radiance ($R(x)$) is computed within the physically plausible boundaries while considering atmospheric light, transmission, and the Radiance Cube constraints.

These equations collectively enforce the Boundary Constraint, contributing to more realistic and physically meaningful results in the context of image dehazing. Adjustments to the parameters (ϵ and λ) may be made based on specific dehazing algorithms and requirements.

The above calculations can be depicted in Fig. 3.

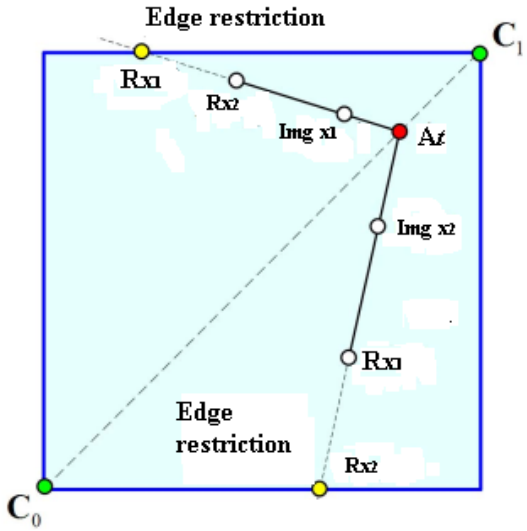


Fig. 3. A cube with boundary limitations that emits light. For every value of x, it is essential that the estimation of $J(x)$ does not surpass the limits of the radiance cube. $J_b(x_1)$ and $J_b(x_2)$ denote the precise positions that define the boundary constraints.

In geometric terms, as described by Eq. 3, when fog affects a pixel $Img(x)$, it tends to shift closer to the overall atmospheric light A_t, as illustrated in Fig. 3. Consequently, this effect can be counteracted by reversing the process and restoring the original clarity of the pixel, denoted as $R(x)$, through a linear extrapolation from A_t to Img_x. The extent of this extrapolation is determined as follows:

$$\frac{1}{st(x)} = \frac{|R(x) - A_t|}{|Img(x) - A_t|} \tag{6}$$

Let us contemplate the notion that the radiance of the scene within a particular image remains constrained; that is to say

$$C_0 \leq R(x) \leq C_1, \quad \forall x \in \Omega \tag{7}$$

In the context of the provided image, two constant vectors, C_0 and C_1, hold significance. Thus, it is imperative that for any given value of *x*, the extrapolation of $Rx)$ should fall within the radiance cube delineated by C_0 and C_1, as depicted in Fig. 3.

The requirement stated above for $R(x)$ leads to an edge restriction on $st(x)$. Assuming the global atmospheric light A_t is provided, the respective boundary constraint point $R(x_1)$ can be calculated for each *x* (see Fig. 3). Consequently, utilizing Eq. 6 and Eq. 7, a minimum value for $st(x)$ can be established, thereby setting the boundary constraint on $st(x)$ as follows:

$$0 \leq st_b(x) \leq st(x) \leq 1 \tag{8}$$

where $st_b(x)$ is the lower bound of $st(x)$, given by

$$st_b(x) = \min \left\{ \max_{c \in \{r,g,b\}} \left(\frac{A_t^c - \text{Img}(x)^c}{A_t^c - C_0^c}, \frac{A_t^c - \text{Img}(x)^c}{A_t^c - C_1^c} \right), 1 \right\} \tag{9}$$

here $Img(x)^c$, ${A_t}^c$, C_0^c and C_1^c are the color channels of $Img(x)$, A_t, C_0 and C_1, respectively

The edge constraint imposed by $st(x)$ provides a novel geometric viewpoint on the well-known dark channel prior [44]. Let's assign $C_0 = 0$ and suppose that the total atmospheric light A_t is brighter than any pixel in the foggy image. This enables us to compute $st_b(x)$ directly using Eq. 3, under the assumption that the dark channel of $R(x)$ at each pixel is zero. In addition, by assuming that the transmission remains constant inside a local image patch, the transmission for each patch $t\ddot{E}IJ(x)$ can be easily calculated using the method described in [44]. This is achieved by applying maximum filtering on $st_b(x)$ using Eq. 10.

$$\tilde{st}(x) = \max_{y \in \omega_x} st_b(y) \tag{10}$$

here ω_x is a local value being originated at *x*.

It is noteworthy that the edge restriction holds greater significance. In general, the ideal global atmospheric light is usually less bright than the brightest pixels in an image. These pixels often come from prominent light sources in the environment, like a brilliant sky or automobile headlights. While the dark channel prior may not effectively capture these bright pixels, the proposed edge restriction remains relevant.

Furthermore, it is of key importance that the commonly utilized constant assumption regarding transmission within a local image patch can be quite stringent. Consequently, the patch-wise transmission, as proposed in [44], [52] based on this assumption, is frequently undervalued. In this context, a more precise patch-wise transmission method is introduced that loosens the aforementioned assumption, thereby permitting slight variations in transmissions within a local patch. Eq. 11 depicts the formation of new patch-wise transformation:

$$\tilde{st}(x) = \min_{y \in \omega_x} \max_{z \in \omega_y} st_b(z) \tag{11}$$

The patch-wise transmission $\tilde{st}(x)$ can be efficiently calculated by simply applying a morphological closing operation on $st_b(x)$. Fig. 4 shows a comparison of the dehazing results achieved by using patch-wise transmissions derived from both the dark channel prior and the boundary constraint map. It is clear that the transmission derived from the dark channel prior is less effective in areas with bright sky. Furthermore, the dehazing results display noticeable halo artifacts.

Fig. 4. Results of image dehazing by applying patch-wise transmission from dark-channel and ERCO-Net (edge restriction). (a) The foggy image, (b) output dehazed image by dark-channel, (c) dehazing result be edge restriction.

A. Contextual Regularization based on L1-norm

Typically, pixels within a localized section of an image exhibit comparable depth values. Leveraging this presumption, a transmission method based on patch-wise constraints has been formulated. Nevertheless, this contextual inference frequently proves inadequate for image sections characterized by sudden depth changes, resulting in pronounced halo artifacts in the dehazing outcomes. This issue can be resolved by computing a weighting function wt on the edge restrictions, given below in Eq. 12.

$$wt(x, y)(st(y) - st(x)) \approx 0 \tag{12}$$

In the context of image processing, the relationship between adjacent pixels, denoted as *x* and *y*, is influenced by a weighting function. This function acts as a sort of "switch" determining the constraint between these pixels. Specifically, when the weighting function, $wt(x, y)$, equals zero, it effectively nullifies the contextual constraint of $st(x)$ between *x* and *y*. The suitable value of $wt(x, y)$ was determined to be inversely proportional to the difference between the values of *x* and *y*; i.e. the larger the distance between *x* and *y*, the lower the weight value. Additionally, it was determined that depth jumps primarily occur at the boundaries of images and within localized areas (or patches), where pixels of similar colors are likely to possess comparable depth values. This suggests that color variation among neighboring pixels can be calculated to create a weighting function. Two instances are provided to

demonstrate the formulation of such weighting functions. One approach entails computing the squared discrepancy between the color vectors of adjacent pixels, as seen in Eq. 13.

$$wt(x,y) = \frac{e^{-|Img(x)-Img(y)|^2}}{2\sigma^2} \tag{13}$$

Here σ acts as tune-able parameter. Same calculation is computed by considering (in Eq. 14) the difference in luminance of the neighboring pixels.

$$wt(x,y) = ((|lum(x) - lum(y)|)^{\alpha} + \epsilon)^{-1} \tag{14}$$

In this context, lum(x) refers to the logarithmic luminance channel of the image Img(x). The exponent $\alpha \geq 0$ determines the level of sensitivity to differences in luminance between two pixels, while ϵ is a small constant (typically 0.0001) used to prevent division by zero. The integration of the weighted contexts throughout the entire image leads to the subsequent contextual optimisation for $st(x)$:

$$\sum_{i \in I} \sum_{j \in \omega_i} wt_{ij} |st_i - st_j| \tag{15}$$

Here *I* is the index set of image pixels, wt_{ij} is the discrete versions of $wt(x,y)$.

B. Estimation of Scene Transmission

In image dehazing, estimating the scene transmission is a crucial step to recover the haze-free image from a hazy input. The scene transmission, denoted by $st(x)$, represents the proportion of light that is transmitted through the atmospheric medium at each pixel x in the image. A common model used for estimating scene transmission is the dark channel prior (DCP) proposed by [44].

The dark channel prior exploits the statistical property that in most outdoor natural images, there exist some pixels with very low intensity values in at least one color channel. This property holds true even in hazy images.

The dark channel $J^{dark}(x)$ of an image J is defined as:

$$J^{dark}(x) = \min_{y \in \Omega(x)} (\min_{c \in \{r,g,b\}} (J^c(y))) \tag{16}$$

where $J^c(y)$ represents the intensity of channel c at pixel y, and $\Omega(x)$ is a local patch centered at pixel x.

The scene transmission $t(x)$ can then be estimated using the dark channel prior as:

$$t(x) = 1 - \omega \cdot \min_{y \in \Omega(x)} (J^{dark}(y)) \tag{17}$$

where ω is a small positive constant used to control the amount of haze removal.

By utilizing the dark channel prior to estimate the scene transmission, the areas of the image impacted by haze can be accurately detected and their intensity appropriately reduced throughout the dehazing procedure.

V. Results with Discussion

Fig. 5 depicts the process of scene transformation estimation of the image performed in different iterations, by applying Eq. 16 and Eq.17.

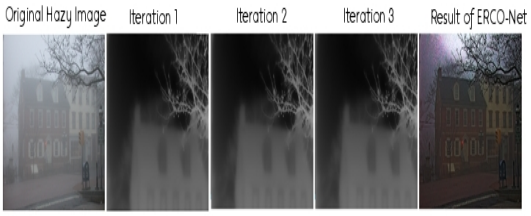


Fig. 5. Results of applying scene transmission estimation on hazy image. The results are shown in iteration followed by the final output.

The results are generated by adjusting the value of *w* to 0.5 in Eq. 17. The edge restriction is set to the C_0 to $(20, 10, 20)^T$ and C_1 to $(150, 150, 300)^T$ in Eq. 9. These adjustments assist in adjusting the haze map with in the view-able window.

The results demonstrate ERCO-Net ability to effectively restore intricate details and vibrant color information within hazy regions of images. It's important to emphasize that the estimated transmissions displayed in the three images on the right side of the figure are not merely scaled versions of depth maps. This is due to the non-uniform distribution of haze across these images, particularly evident in image featuring expansive clear sky areas. Essentially, the transmission function serves as a reflection of the haze density present within the captured image, rather than simply representing depth.

A. Comparison with Other Approaches

For evaluation purposes, the ERCO-Net results were also compared with some state-of-the-art models. These models include dark channel prior (DCP) [44], Atmospheric Light Estimation (ALT) [53], Image Fusion [54], and multiscale retinex [55].

Fig. 6 depicts the comparison of our approach with the algorithms aforementioned.

Fig. 6. Comparison with DCP, ALT, Image Fusion, and MSR approaches using ERCO-Net on the NH-Haze2 dataset [56].

Similarly, Fig. 7 shows the output of our algorithm in comparison with other models.

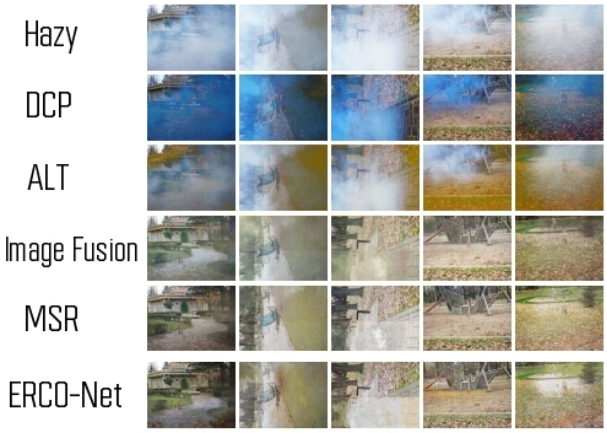


Fig. 7. Comparison with DCP, ALT, Image Fusion, and MSR approaches using ERCO-Net on the NH-Haze2 dataset [57].

For evaluation purposes, two metrics are employed: the Structural Similarity Index (SSIM) and the Peak Signal-to-Noise Ratio (PSNR). These metrics are frequently utilized in image processing to evaluate the quality of images.

Table II presents the comparison of our approach with other algorithms over NHHaze and NHHaze-2 datasets.

TABLE II. QUANTITATIVE COMPARISONS OVER NHHAZE AND NHHAZE-2 DATASETS FOR DIFFERENT METHODS AND ERCO-NET APPROACH

Methods	NHHaze		NHHaze2	
	PSNR	SSIM	PSNR	SSIM
DCP	16.62	0.817	12.92	0.505
ALT	19.06	0.85	17.69	0.616
Image fusion	30.23	0.975	19.5	0.66
MSR	34.59	0.975	23.53	0.754
ERCO-Net	**37.61**	**0.991**	**25.54**	**0.783**

VI. CONCLUSION

In the field of single image dehazing, a significant challenge arises from the inherent ambiguity between image color and depth. This ambiguity often leads to difficulty in distinguishing between haze-affected and unaffected pixels, particularly when their coloration is similar. ERCO-Net addresses this challenge by combining edge restriction and contextual optimization, which improves image clarity and preserves critical details. The method demonstrated reliable dehazing performance with reduced computational complexity, highlighting its potential for practical applications. Future work will focus on further optimizing ERCO-Net for large-scale datasets and addressing limitations in extreme haze conditions. Additionally, improvements in edge detail retention and the exploration of its applicability to other image restoration tasks will be key areas of investigation.

ACKNOWLEDGMENT

We would like to express sincere gratitude to AlMaarefa University, Riyadh, Saudi Arabia, for supporting this research.

DECLARATIONS

Conflicts of Interests / Competing Interests

The authors declare that they have no conflict of interest.

REFERENCES

[1] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2018.

[2] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma, "Contrastive learning for compact single image dehazing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10 551–10 560.

[3] L. Li, Y. Dong, W. Ren, J. Pan, C. Gao, N. Sang, and M.-H. Yang, "Semi-supervised image dehazing," *IEEE Transactions on Image Processing*, vol. 29, pp. 2766–2779, 2019.

[4] A. Golts, D. Freedman, and M. Elad, "Unsupervised single image dehazing using dark channel prior loss," *IEEE transactions on Image Processing*, vol. 29, pp. 2692–2701, 2019.

[5] Z. Zhu, Y. Luo, H. Wei, Y. Li, G. Qi, N. Mazur, Y. Li, and P. Li, "Atmospheric light estimation based remote sensing image dehazing," *Remote Sensing*, vol. 13, no. 13, p. 2432, 2021.

[6] S. C. Agrawal and A. S. Jalal, "A comprehensive review on analysis and implementation of recent image dehazing methods," *Archives of Computational Methods in Engineering*, vol. 29, no. 7, pp. 4799–4850, 2022.

[7] Y. Song, Z. He, H. Qian, and X. Du, "Vision transformers for single image dehazing," *IEEE Transactions on Image Processing*, vol. 32, pp. 1927–1941, 2023.

[8] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4770–4778.

[9] W. Ren, J. Pan, H. Zhang, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks with holistic edges," *International Journal of Computer Vision*, vol. 128, pp. 240–259, 2020.

[10] P. Sharma, P. Jain, and A. Sur, "Scale-aware conditional generative adversarial network for image dehazing," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 2355–2365.

[11] X. Zhang, T. Wang, J. Wang, G. Tang, and L. Zhao, "Pyramid channel-based feature attention network for image dehazing," *Computer Vision and Image Understanding*, vol. 197, p. 103003, 2020.

[12] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua, "Gated context aggregation network for image dehazing and deraining," in *2019 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2019, pp. 1375–1383.

[13] Y.-W. Lee, L.-K. Wong, and J. See, "Image dehazing with contextualized attentive u-net," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 1068–1072.

[14] L. Yan, W. Zheng, C. Gou, and F.-Y. Wang, "Feature aggregation attention network for single image dehazing," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 923–927.

[15] S. Zhang and F. He, "Drcdn: learning deep residual convolutional dehazing networks," *The Visual Computer*, vol. 36, no. 9, pp. 1797–1808, 2020.

[16] S. Zhang, F. He, and W. Ren, "Photo-realistic dehazing via contextual generative adversarial networks," *Machine Vision and Applications*, vol. 31, no. 5, p. 33, 2020.

[17] A. Dudhane, H. Singh Aulakh, and S. Murala, "Ri-gan: An end-to-end network for single image haze removal," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2019, pp. 0–0.

[18] G. Tang, L. Zhao, R. Jiang, and X. Zhang, "Single image dehazing via lightweight multi-scale networks," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 5062–5069.

[19] A. Wang, W. Wang, J. Liu, and N. Gu, "Aipnet: Image-to-image single image dehazing with atmospheric illumination prior," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 381–393, 2018.

[20] C. Wang, Y. Zou, and Z. Chen, "Abc-net: Avoiding blocking effect & color shift network for single image dehazing via restraining transmission bias," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 1053–1057.

[21] S. Chen, Y. Chen, Y. Qu, J. Huang, and M. Hong, "Multi-scale adaptive dehazing network," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2019, pp. 0–0.

[22] A. Singh, A. Bhave, and D. K. Prasad, "Single image dehazing for a variety of haze scenarios using back projected pyramid network," in *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16*. Springer, 2020, pp. 166–181.

[23] S. Yin, Y. Wang, and Y.-H. Yang, "A novel image-dehazing network with a parallel attention block," *Pattern Recognition*, vol. 102, p. 107255, 2020.

[24] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3194–3203.

[25] H. Zhang, V. Sindagi, and V. M. Patel, "Multi-scale single image dehazing using perceptual pyramid deep network," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 902–911.

[26] D. Zhao, L. Xu, L. Ma, J. Li, and Y. Yan, "Pyramid global context network for image dehazing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 8, pp. 3037–3050, 2020.

[27] C.-L. Guo, Q. Yan, S. Anwar, R. Cong, W. Ren, and C. Li, "Image dehazing transformer with transmission-aware 3d position embedding," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5812–5820.

[28] X. Chen, H. Lu, K. Cheng, Y. Ma, Q. Zhou, and Y. Zhao, "Sequentially refined spatial and channel-wise feature aggregation in encoder-decoder network for single image dehazing," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 2776–2780.

[29] M. Hong, Y. Xie, C. Li, and Y. Qu, "Distilling image dehazing with heterogeneous task imitation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 3462–3471.

[30] X. Liang, R. Li, and J. Tang, "Selective attention network for image dehazing and deraining," in *Proceedings of the 1st ACM International Conference on Multimedia in Asia*, 2019, pp. 1–6.

[31] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced pix2pix dehazing network," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 8160–8168.

[32] Z. Zhang, L. Zhao, Y. Liu, S. Zhang, and J. Yang, "Unified density-aware image dehazing and object detection in real-world hazy scenes," in *Proceedings of the Asian Conference on Computer Vision*, 2020.

[33] R. Chen and E. M.-K. Lai, "Convolutional autoencoder for single image dehazing." in *ICIP*, 2019, pp. 4464–4468.

[34] Y. Dong, Y. Liu, H. Zhang, S. Chen, and Y. Qiao, "Fd-gan: Generative adversarial networks with fusion-discriminator for single image dehazing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 10 729–10 736.

[35] H. Zhu, X. Peng, V. Chandrasekhar, L. Li, and J.-H. Lim, "Dehazegan: When image dehazing meets differential programming." in *IJCAI*, 2018, pp. 1234–1240.

[36] F. Yang and Q. Zhang, "Depth aware image dehazing," *The Visual Computer*, vol. 38, no. 5, pp. 1579–1587, 2022.

[37] H. Dong, X. Zhang, Y. Guo, and F. Wang, "Deep multi-scale gabor wavelet network for image restoration," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 2028–2032.

[38] X. Liu, Y. Ma, Z. Shi, and J. Chen, "Griddehazenet: Attention-based multi-scale network for image dehazing," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 7314–7323.

[39] K. Metwaly, X. Li, T. Guo, and V. Monga, "Nonlocal channel attention for nonhomogeneous image dehazing," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 452–453.

[40] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 11 908–11 915.

[41] J. Wang, C. Li, and S. Xu, "An ensemble multi-scale residual attention network (emra-net) for image dehazing," *Multimedia Tools and Applications*, vol. 80, no. 19, pp. 29 299–29 319, 2021.

[42] S. Yin, X. Yang, Y. Wang, and Y.-H. Yang, "Visual attention dehazing network with multi-level features refinement and fusion," *Pattern Recognition*, vol. 118, p. 108021, 2021.

[43] P. Shyam and H. Yoo, "Data efficient single image dehazing via adversarial auto-augmentation and extended atmospheric scattering model," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 227–237.

[44] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.

[45] L. Chen, W. Wu, C. Fu, X. Han, and Y. Zhang, "Weakly supervised semantic segmentation with boundary exploration," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVI 16*. Springer, 2020, pp. 347–362.

[46] Y. Liu, J. Pan, J. Ren, and Z. Su, "Learning deep priors for image dehazing," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 2492–2500.

[47] J. Wan, Z. Lai, J. Li, J. Zhou, and C. Gao, "Robust facial landmark detection by multiorder multiconstraint deep networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 5, pp. 2181–2194, 2021.

[48] D. Yang and J. Sun, "Proximal dehaze-net: A prior learning-based deep network for single image dehazing," in *Proceedings of the european conference on computer vision (ECCV)*, 2018, pp. 702–717.

[49] R. Fattal, "Single image dehazing," *ACM transactions on graphics (TOG)*, vol. 27, no. 3, pp. 1–9, 2008.

[50] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE transactions on pattern analysis and machine intelligence*, vol. 25, no. 6, pp. 713–724, 2003.

[51] R. T. Tan, "Visibility in bad weather from a single image," in *2008 IEEE conference on computer vision and pattern recognition*. IEEE, 2008, pp. 1–8.

[52] W.-C. Cheng, H.-C. Hsiao, W.-L. Huang, and C.-H. Hsieh, "Image haze removal using dark channel prior technology with adaptive mask size." *Sensors & Materials*, vol. 32, 2020.

[53] H. Lu, Y. Li, S. Nakashima, and S. Serikawa, "Single image dehazing through improved atmospheric light estimation," *Multimedia Tools and Applications*, vol. 75, pp. 17 081–17 096, 2016.

[54] Z. Zhu, H. Wei, G. Hu, Y. Li, G. Qi, and N. Mazur, "A novel fast single image dehazing algorithm based on artificial multiexposure image fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–23, 2020.

[55] J. Wang, K. Lu, J. Xue, N. He, and L. Shao, "Single image dehazing based on the physical model and msrcr algorithm," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2190–2199, 2017.

[56] C. O. Ancuti, C. Ancuti, F.-A. Vasluianu, and R. Timofte, "Ntire 2021 nonhomogeneous dehazing challenge report," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 627–646.

[57] C. O. Ancuti, C. Ancuti, and R. Timofte, "Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 444–445.

Optimizing Text Summarization with Sentence Clustering and Natural Language Processing

Zahir Edress*, Yasin Ortakci
Department of Computer Engineering, Karabuk University, Karabuk, Turkey

***Abstract*—Text summarization is an important task in natural language processing (NLP), with significant implications for information retrieval and content management. Traditional summarization methods often struggle with issues like redundancy, loss of key information, and inability to capture the underlying semantic structure of the text. This paper addresses these challenges by presenting an advanced approach to extractive summarization, which integrates clustering-based sentence selection with the BART model.The proposed method tackles the problem of redundancy by using Term Frequency-Inverse Document Frequency (TF-IDF) for feature extraction, followed by K-means clustering to group similar sentences. This clustering step is designed to reduce redundancy by ensuring that each cluster represents a distinct theme or topic. Representative sentences are then selected from these clusters based on their cosine similarity to a user query, which helps in retaining the most relevant information. These selected sentences are then fed into the BART model to generate the final abstractive summary. This combination of extractive and abstractive techniques addresses the common problem of information loss, ensuring that the summary is both comprehensive and coherent. The approach is evaluated using the CNN/DailyMail and XSum datasets, which are widely recognized benchmarks in the summarization domain. Results assessed through ROUGE metrics demonstrate that the proposed model substantially improves summarization quality compared to existing benchmarks.**

Keywords*—Abstractive summarization; extractive summarization; sentence clustering; language understanding; information retrieval

I. Introduction

Text summarization task aims to generate a concise and coherent summary from a given text while preserving its key information and meaning[1]. The challenge lies in effectively capturing the essence of the original content, which can vary widely in length, complexity, and structure. In the realm of text summarization, there are two primary approaches [2]: extractive and abstractive summarization, extractive summarization involves selecting and compiling sentences or phrases directly from the source text to form a summary. This approach relies on identifying the most important segments of the text, but may lack coherence and fluidity in the final output. On the other hand, abstractive summarization involves generating new sentences that convey the core ideas of the text [3], often resulting in more human-like and fluent summaries. However, this approach requires advanced techniques to ensure that the generated summary is both accurate and contextually relevant [4].

Extractive summarization is the most widely utilized and rapidly developed approach [5]. Despite its advancements, extractive summarization faces challenges, particularly in initial center selection and redundancy issues. In articles with complex sentence structures, the summarized sentences often exhibit high redundancy. This occurs because the extracted sentences may share significant semantic similarity, leading to repetitive content in the summary. Consequently, even if the summary includes relevant concepts, it may contain redundant information due to the repetition of similar concepts. This redundancy can make the summary unnecessarily lengthy and less concise. To solve the above problems this paper explores an enhanced summarization approach that combines clustering-based extractive methods with the BART model. the methodology starts by employing TF-IDF to evaluate the significance of each sentence within the original text and subsequently apply K-means clustering to group similar sentences into clusters. This process helps in identifying thematic groups within the text, which reduces redundancy and ensures a more diverse representation of content. For each cluster, representative sentences are selected based on cosine similarity to a query sentence. By selecting sentences that closely match this query, key information from each cluster is accurately captured, the selected sentences are then used to form a context for the BART model, which generates the final summary.

The proposed approach is evaluated using several well-established benchmark datasets, including CNN/DailyMail and XSum. Performance is assessed using ROUGE metrics, and the experimental results indicate that the model achieves superior ROUGE scores compared to other benchmark summarization methods. Specifically, the approach demonstrates significant improvements in terms of informativeness and coherence, reflecting a more accurate and fluid representation of the original text.

The main contributions of this work are summarized as follows:

- A novel framework for text summarization is proposed that enhances the coherence and human-likeness of the generated summaries.
- An effective sentence sampling strategy is introduced within the rewrite model, which significantly improves the quality and relevance of the summaries.
- The extensive experiments demonstrate that the proposed model consistently outperforms existing state-of-the-art baselines in text summarization.

The rest of the paper is organized as follows. Section II reviews related work of this paper. Section III gives a detailed explanation of the methodology. Section IV describes the experimental setups, which is followed by the experimental results in Section V. The paper is concluded in Section VI.

*Corresponding author

II. Related Work

Automatic Text summarization (ATS) has been extensively researched, with various methods proposed to address its challenges [6]. Early work in extractive summarization focused on approaches such as frequency-based methods, graph-based algorithms like Text Rank, and machine learning techniques [7]. These methods generally perform well in selecting salient sentences but often struggle with generating coherent summaries. Recent advancements in abstractive summarization have been driven by transformer-based models. Models like BERT (Bidirectional Encoder Representations from Transformers) and its variants have achieved state-of-the-art performance on summarization tasks by capturing deep contextual information [8]. BART a more recent model, combines the strengths of both bidirectional and autoregressive transformers, making it particularly effective for text generation tasks [1]. This section provides a summary of relevant research that informs the hybrid summarization approach, which combines extractive and abstractive techniques.

A. Extractive Text Summarization

The extractive summarization method involves selecting key sentences and keywords directly from the original text [9], It works by scoring sentences based on their importance, with higher scores indicating more significant sentences [10]. The summary is created by sequentially choosing the highest-scoring sentences from the text [11]. Liu et al. [7] enhanced extractive summarization by combining the improved TextRank algorithm with K-means clustering, This approach utilizes the BM25 model to compute sentence similarity and derive TR scores which are used to select initial cluster centers and reduce redundancy. Mohsen et al. [12] developed a hierarchical self-attentive neural network model that integrates reinforcement and supervised learning to rank sentences by directly optimizing the ROUGE metric. The model leverages hierarchical self-attention to create document and sentence embeddings that capture the document's structure, enhancing feature representation and summarization quality.

B. Abstractive Text Summarization

The abstractive summarization method creates summaries by generating new words and phrases, rather than directly extracting from the original text [13]. It uses natural language understanding to analyze the grammar and semantics of the document [14], allowing it to convey the main ideas in a newly formulated text [15]. Jain et al. [16] developed a specialized dataset for abstractive summarization to increase the number of training samples while maintaining manageable lengths for each sample. Bahrainian et al. [17] developed CATS an advanced abstractive neural summarization model that enhances the traditional sequence-to-sequence approach by introducing a mechanism for controlling the latent topic distribution of the generated summaries Su et al. [18] proposed a two-stage method for variable-length abstractive summarization using a text segmentation module and a Transformer-based summarization model. It first segments the text using BERT and BiLSTM, then applies a two-stage training approach with BERTSUM for document and segment summarization.

C. Hybrid Text Summarization

The hybrid summarization method combines both extractive and abstractive techniques to leverage the strengths of each approach. Initially, an extractive summary is generated using extraction models or trainers [8]. This summary is then refined through semantic analysis and rephrasing to produce an abstractive summary [19]. This dual approach aims to create a more comprehensive summary by integrating the benefits of both methods [20]. Recent studies have increasingly focused on combining extractive and abstractive methods [21]. For instance, using extractive summaries as inputs for abstractive models has demonstrated improved performance [22], highlighting the efficacy of this integrated approach in enhancing summary quality. Zhang et al. [1] proposed framework as Extract-then-Abstract approach, an initial extractive summarization model is used to identify and select relevant sentences from the text. This extractive summary is then refined using an abstractive model to generate summary. Morozovskii and Ramanna [23] proposed supervised learning model employs a hybrid approach that integrates both extractive and abstractive elements to enhance the inclusion of crucial information in news summaries. Wang et al. [13] introduced the Topic-Injected Bidirectional Encoder Representations from Transformers (TP-BERT) by incorporates contrastive learning during training.

The previous studies highlights various methodologies and advancements in text summarization. Table I summarizes different studies, detailing the year of the study, the methods employed, key contributions, and the limitations encountered. the proposed approach introduces a novel enhancement by integrating clustering-based sentence selection with the BART model. Unlike traditional extractive methods that select sentences based on frequency or similarity alone, the method utilizes TF-IDF for feature extraction and K-means clustering to group similar sentences. This clustering process identifies thematic groups within the text, reducing redundancy and ensuring a more diverse representation of content. By selecting representative sentences from each cluster based on cosine similarity to a query sentence, the approach effectively captures essential information while minimizing redundancy. This method addresses the limitations of improved TextRank, which may not fully achieve thematic coherence. Moreover, the integration with BART enhances the fluency and coherence of the final summary, leading to more contextually accurate and readable outputs.

III. Methodology

Our proposed summarization approach integrates extractive and abstractive techniques to enhance the relevance and coherence of generated summaries. The methodology is designed to capture key information from the source text while producing fluent and contextually accurate summaries. The process involves data preprocessing, vectorization, clustering-based sentence selection, representative sentence extraction, and abstractive summarization using BART as shown in Fig. 1.

A. Data Preprocessing

The preprocessing steps included Articles were retrieved from the dataset based on user-defined keywords using a

TABLE I. SUMMARY OF RELATED WORK IN TEXT SUMMARIZATION

Study	Year	Method	Key Contributions	Limitations
Wang et al [13]	2024	Integrates topic words into sentences and uses contrastive learning during training	Improving semantic consistency and summarization quality	Incorporating topic words and contrastive learning may add complexity to the model, potentially increasing computational requirements.
Jain et al [24]	2024	Generating additional training samples by creating multiple extractive summaries	Handling long documents	Potential information loss from segmenting documents into 512-token
Liu et al [7]	2024	Employs an improved BM25 model to compute BM25 similarity between sentences.	Reduce sentence redundancy	The method does not account for the position of sentences in the final summary, which might affect the coherence and readability of the summary
Morozovskii and Ramanna [23]	2023	Proposes an enhancement to the transformer model's attention mechanism by integrating frequency information for each word to better handle rare words.	Enhanced Attention Mechanism	The model may include less relevant information due to its distribution of attention across words
Zhang et al [1]	2023	Implements sentence sampling strategy to create parallel data for the rewrite model without manual annotation	Develops a pipeline using topic modeling to select the most relevant sentences	The extractive stage might still result in summaries that lack smooth transitions between sentences
Bahrainian et al[17]	2022	Neural sequence-to-sequence model with encoder-decoder and topical attention	Introduces customizable abstractive summarization using topic modeling.	Need for comparison with transformer-based models.
Su et al [18]	2020	Two-stage Transformer-based summarization model with text segmentation	Employs BERT and LSTM for segmentation and uses collaborative training.	The integrating of multiple models and collaborative training might imply potential challenges in practical implementation.
Mohsen et al [12]	2020	Hierarchical self-attentive reinforced neural network-based summarization model	Combines reinforcement and supervised learning to optimize the ROUGE evaluation metric directly.	The complexity of the model may lead to increased computational requirements.

search function that scans through the dataset and returns the top results. This search is limited to a maximum of five articles to maintain manageability. Each article was segmented into sentences using NLTK's "sent tokenize" function to facilitate clustering and similarity computations [25]. Sentence segmentation enables us to process and analyze each sentence independently [26]. Further preprocessing included stop words removal, which involved eliminating common but less informative words (e.g. "the", "and") to focus on more meaningful terms [27]. Also, word alignment process is to remove punctuation and numbers from each sentence. Finally, Lemmatization was applied to reduce words to their base or root forms (e.g. transforming "running" to "run") [28], Fig. 2 summarizes these steps.

B. Vectorization

In this step, the Term Frequency-Inverse Document Frequency (TF-IDF) vectorizer is employed to convert sentences into numerical feature vectors, this process begins with applying the vectorizer to the preprocessed sentences, transforming them into a high-dimensional space where each dimension corresponds to a term in the vocabulary. The (TF-IDF) model assigns weights to terms based on their importance within each sentence relative to the entire corpus. By capturing both the frequency of terms within individual sentences (Term Frequency) and their significance across the corpus, the TF-IDF vectorizer produces feature vectors that effectively represent the content and context of each sentence.

C. Clustering Process

A clustering-based approach was employed to identify representative sentences [7]. Specifically, the K-means algorithm was used to cluster sentences into a predefined number of clusters. K-means partitions sentences based on their vector representations [1], grouping similar sentences together [29]. the elbow method was utilized to determine the optimal number of clusters as illustrated in Fig. 3. This method involves plotting the within-cluster sum of squares (inertia) against the number of clusters [30]. Inertia measures the sum of squared distances between each point and its assigned cluster center, reflecting the compactness of the clusters [31]. the K-means algorithm was executed across a range of cluster counts (from 1 to 10) and computed the inertia for each count.The resulting plot displays the number of clusters on the x-axis and the corresponding inertia on the y-axis. The "elbow" of the plot, where the rate of decrease in inertia slows down, indicates the optimal number of clusters. This point represents a balance between minimizing inertia and avoiding overfitting by selecting too many clusters [32]. choosing the optimal number of clusters affects the thematic diversity and coherence of the summary. Fewer clusters may lead to excessive redundancy, while more clusters can cause over-segmentation and loss of crucial information. The elbow method helps in finding a balance that improves summary quality by ensuring each

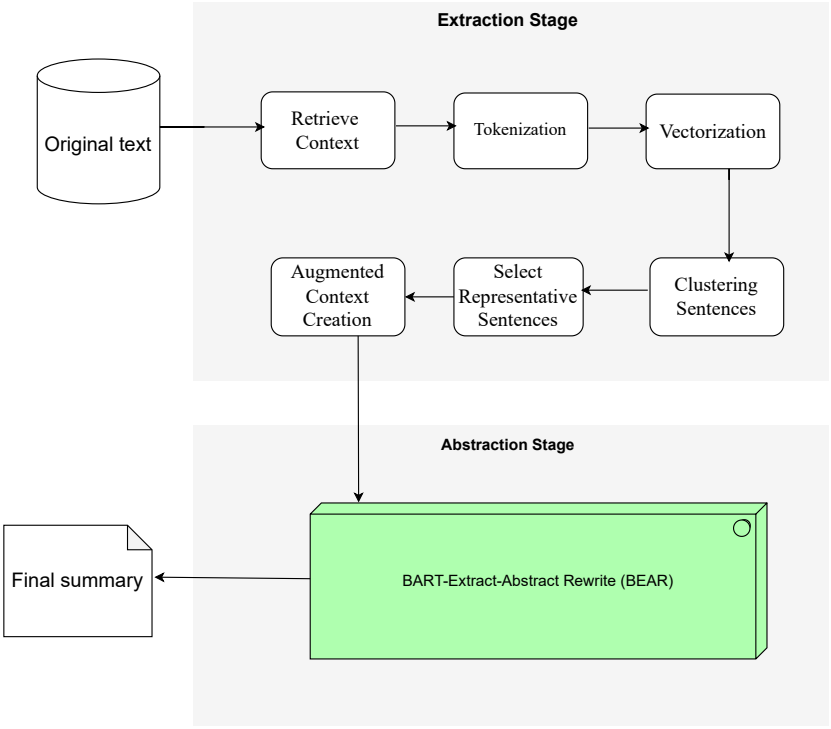


Fig. 1. The general structure of the proposed model.

cluster represents a distinct theme or topic, which enhances the effectiveness and the summarization results [33].

D. Selecting Representative Sentences

To refine the selection of representative sentences from each cluster, cosine similarity was calculated between a query sentence and each sentence within a cluster. The query was constructed from the first few sentences of the article to effectively represent the main context. For each cluster, the sentence with the highest cosine similarity to the query was selected as outlined in the Algorithm 1. cosine similarity was chosen because it effectively quantifies the similarity between sentences based on their term frequency representation [34], this metric is particularly useful for identifying sentences that closely match the thematic context of the query sentence [35], ensuring that the representative sentences are contextually relevant.

E. Summarization with BART

The BART (Bidirectional and Auto-Regressive Transformers) model, specifically the pre-trained facebook/bart-large-cnn version was used for generating summaries due to its strong performance on summarization tasks [1]. The model was configured to produce summaries with a maximum length of 150 tokens and a minimum of 40 tokens. It was fed with an augmented context, which was created by concatenating representative sentences selected through a clustering process. This augmented context helped the model generate a more coherent and comprehensive final summary.

Algorithm 1: Clustering-Based Extractive-Then-Abstractive Summarization

Input Text document D, User query Q
Output Final summary S
Feature Extraction: Calculate the Term Frequency-Inverse Document Frequency (TF-IDF) for each sentence in document D Represent each sentence as a TF-IDF vector
Sentence Clustering: Apply K-means clustering to group the sentences into K clusters based on their TF-IDF vectors
Representative Sentence Selection: for *each cluster* **do**
Calculate the cosine similarity between each sentence and the user query Q Select the sentence with the highest cosine similarity within each cluster as the representative sentence
Input to BART Model: Combine the selected representative sentences into a coherent sequence Feed the sequence of selected sentences into the pre-trained BART model
Generate Final Summary: The BART model generates an abstractive summary S from the input sequence of representative sentences
return Final summary S

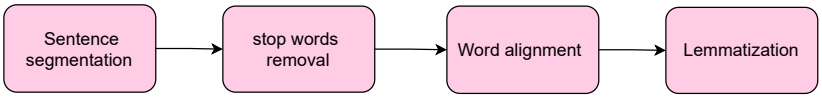


Fig. 2. Data preprocessing steps.

IV. Experiments

In this section, The experiments conducted to evaluate the effectiveness of the proposed summarization model are presented. These experiments are designed to assess the performance of various summarization methods across different datasets, with a focus on comparing baseline methods and the proposed approach.

A. Datasets

To evaluate the performance of the framework, it is assessed using two benchmark summarization datasets: CNN/DailyMail [36] and XSum [37], detailed statistical information for these datasets is provided in Table II.

1) The CNN/DailyMail: dataset is widely used for evaluating text summarization models. It consists of approximately 93,000 articles from CNN and 220,000 articles from the Daily Mail newspapers, totaling around 313,000 news articles [6]. Each article in the dataset is paired with multiple reference summaries, which are generally several sentences long, and are written by humans. The summaries are derived from the article highlights, typically found as bullet points making this dataset suitable for both extractive and abstractive summarization tasks.

2) The XSum: (Extreme Summarization) dataset is a benchmark dataset specifically designed for single-document abstractive summarization tasks. It consists of approximately 227,000 online articles from the BBC, covering a wide range of topics, including news, sports, and lifestyle. Each article is paired with a corresponding one-sentence summary, known as the "extreme summary", which captures the key point of the article. The one-sentence summaries are often highly abstractive, meaning they are not simply extracted sentences from the original text but are instead newly generated sentences that convey the main idea [7].

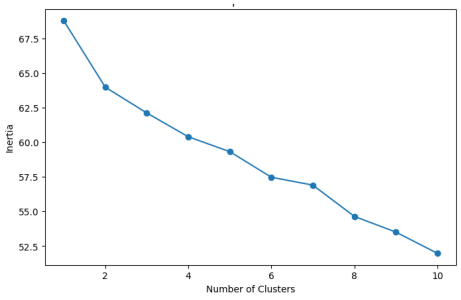


Fig. 3. Elbow method for determining the optimal number of clusters.

B. Baseline Methods

A range of high-performing models is selected for evaluation, relevant models from recent years as the baseline.

TABLE II. Statistics of the CNN/DailyMail and XSum Datasets

Dataset	#Docs	#Avg. sent.	#Avg. words	Domain
CNN/DailyMail	287,113	45.7	781	News
XSum	204,045	19.77	431	News

1) Lead-3: method simply selects the first three sentences from the document as the summary. It exploits the tendency of news articles to place the most important information at the beginning [38].

2) PacSum: (Parameterized Summarization) uses a heuristic-based approach to select sentences based on their importance, as determined by a parameterized model. It is known for its effectiveness in extracting informative sentences that contribute to high-quality summaries [39].
FAR (Focus-Aware Ranking) incorporates focus-aware ranking strategies to select the most relevant sentences. It improves upon simple extraction methods by considering the relevance and focus of the content [40].

3) SUMO: (Summarization with Meaningful Overlaps) focuses on identifying and using significant overlaps in sentence content to improve summary quality. It aims to balance informativeness and redundancy in the selected sentences [41].

4) PGNet: (Position-Aware Graph Network) incorporates position information and graph-based methods to rank sentences based on their importance in a given document. It improves upon traditional methods by leveraging positional information [16].

5) REFRESH: (REsponse-aware Frequency-based Summarization with Contextualized Highlights) uses response-aware frequency metrics to highlight key content. It adapts based on context to improve summary relevance [42].

6) SEQ: (Sequence-to-Sequence) models are based on sequence-to-sequence learning approaches that generate summaries by learning from large amounts of text. These models often use encoder-decoder architectures to produce coherent and contextually relevant summaries [43].
TED (Text Extensible Domain) leverages domain-specific knowledge to enhance summarization. By extending the text domain, TED models improve summary quality by incorporating additional contextual information [44].

7) CPSUM: (Soft) is an extractive summarization approach that employs a soft selection mechanism, focusing on selecting important sentences based on their relevance and contribution to the summary [13].

8) CPSUM: (Hard) differs from its soft counterpart by using a hard selection mechanism, which selects sentences more decisively without soft probability weighting [45].

9) BERTEXT: is an extractive summarization method that employs sequence labeling. This approach determines which sentences should be included in the summary by leveraging predictions made by a pre-trained BERT model [8].

10) BART: effectively captures context through its bidirectional encoder and produces high-quality summaries using its autoregressive decoder [1].

C. Evaluation

To evaluate the performance of the summarization approach, the ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metric was used to assess the quality of the generated summaries [19]. ROUGE measures the overlap between the generated summaries and reference summaries across several dimensions including [46]:

ROUGE-1 (R-1): Measures unigram overlap between the generated and reference summaries. This metric evaluates the proportion of unigrams (single words) in the generated summary that are present in the reference summaries [4].

ROUGE-2 (R-2): Assesses bigram overlap, focusing on pairs of consecutive words. This metric helps determine how well the generated summary captures the sequence of word pairs found in the reference summaries [13].

ROUGE-L (R-L): Evaluates the longest common subsequence between the generated and reference summaries [47]. The calculation of the ROUGE-N is shown in Eq. (1).

$$\text{ROUGE-N} = \frac{\sum_{s \in S} \sum_{n \in \text{N-grams}(s)} \text{count}_{\text{match}}(n)}{\sum_{s \in S} \sum_{n \in \text{N-grams}(s)} \text{count}_{\text{ref}}(n)} \quad (1)$$

where:

- S is the set of generated summaries.
- N-grams(s) represents the set of n-grams in summary s.
- $\text{count}_{\text{match}}(n)$ is the count of n-grams n that match between the generated summary and the reference summaries.
- $\text{count}_{\text{ref}}(n)$ is the count of n-grams n in the reference summaries.

D. Implementation

The implementation involved Libraries, the "transformers" library for BART was used, "datasets" for loading the datasets, "sklearn" for clustering and vectorization, and "nltk" for sentence tokenization. A Python script was developed to automate the entire process, from article retrieval and sentence clustering to summary generation and evaluation.

V. Results and Discussion

A. Results

The performance of the proposed BEAR model is evaluated using ROUGE metrics across various summarization models, as detailed in Table III. This table showcases the R-1, R-2, and R-L scores for different models on the CNN/DailyMail dataset, including the BEAR model. Notably, BEAR demonstrates competitive performance with R-1, R-2, and R-L scores of 48.15, 19.23, and 37.03, respectively, surpassing several baseline models.

To facilitate a visual interpretation of these results, Fig. 4 presents a heatmap of the ROUGE scores. This heatmap highlights the comparative performance of BEAR against other models, clearly illustrating that BEAR excels in both R-1 and R-2 metrics. This indicates its superior capability in capturing relevant information and ensuring coherence in summaries.

TABLE III. ROUGE Scores for Different Models on the CNN/DailyMail Dataset

Model	R-1	R-2	R-L
Lead 3	40.01	17.45	36.31
PacSum	40.37	17.92	36.62
FAR	40.42	17.95	36.67
SUMO	41.00	18.40	37.20
PGNet	39.50	17.30	36.40
REFRESH	41.30	18.40	37.50
SEQ	23.24	7.10	22.15
TED	38.73	16.84	36.15
BART	44.16	18.07	35.53
BEAR (proposed model)	**48.15**	**19.23**	**37.03**

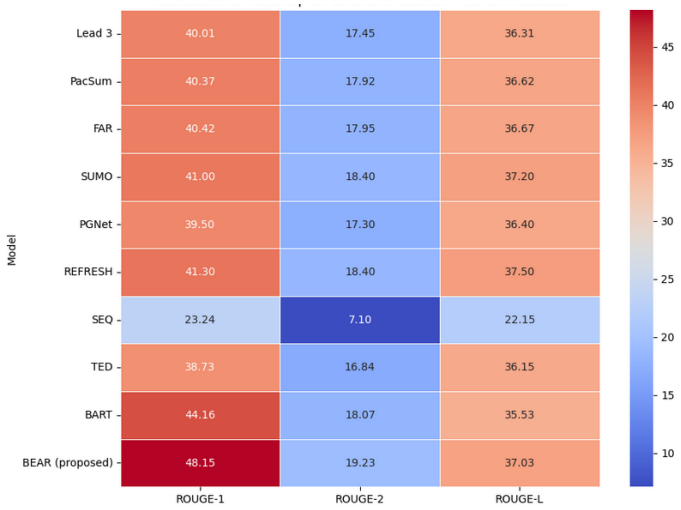


Fig. 4. Heatmap of ROUGE scores for different summarization models on the CNN/DailyMail dataset.

Table IV presents the ROUGE scores for various summarization models evaluated on the XSum dataset. Notably, the proposed BEAR model demonstrates impressive performance, achieving R-1, R-2, and R-L scores of 25.00, 8.57, and 19.44, respectively. Fig. 5 provides a visual representation of these results through a heatmap. In the heatmap, color gradients are used to illustrate the relative performance of each model, with darker shades indicating higher ROUGE scores. This visualization effectively highlights the strengths of the BEAR model compared to other models, underscoring its capability in generating high-quality summaries.

TABLE IV. ROUGE SCORES FOR VARIOUS SUMMARIZATION MODELS ON THE XSUM DATASET

Model	R-1	R-2	R-L
Lead 3	16.30	1.60	11.95
SEQ	20.11	5.23	16.15
CPSUM(Soft)	17.22	2.17	12.71
CPSUM(Hard)	17.29	2.18	12.73
BERTEXT	22.86	4.48	17.16
BEAR (proposed model)	**25.00**	**8.57**	**19.44**

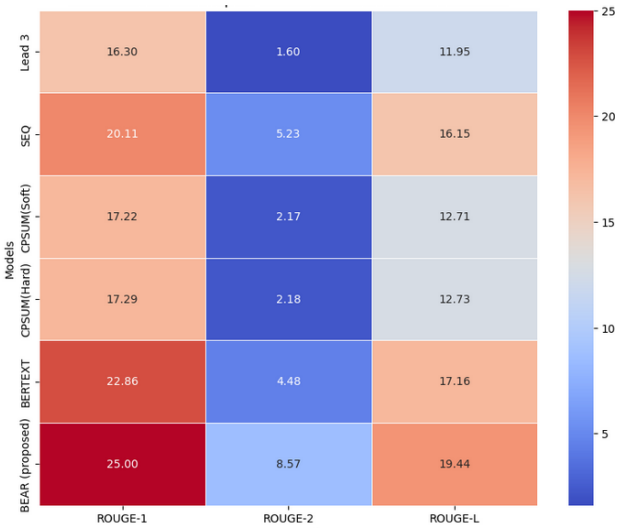


Fig. 5. Heatmap of ROUGE scores for various summarization models on the XSum dataset.

B. Comparative Analysis with Previous Research

To assess the relevance and competitiveness of the BEAR model, its results were compared with those of other state-of-the-art summarization approaches. Table V shows a detailed comparison of ROUGE scores across different models on the CNN/DailyMail and XSUM datasets, highlighting the BEAR model's superior performance.

1) CNN/DailyMail dataset: The BEAR model outperforms several baseline models such as Lead-3, PacSum, and REFRESH in terms of R-1 and R-2 scores, indicating its superior ability to capture essential content. For example, Lead-3 achieved an R-1 score of 40.01, while BEAR obtained 48.15, demonstrating significant improvements compared to BERTSUM, which has been widely used for summarization, BEAR's higher ROUGE scores underscore its effectiveness in integrating clustering-based sentence selection with abstractive summarization.

2) XSum dataset: The BEAR model outperforms models like BERTEXT and SEQ in both R-1 and R-2 scores. The comparative analysis of the proposed BEAR model against other summarization models on the XSum dataset demonstrates its superior performance across all three ROUGE metrics (R-1, R-2, and R-L) key observations include:

3) ROUGE-1 (R-1): BEAR achieved an R-1 score of 25.00, outperforming other models, such as SEQ (20.11) and BERTEXT (22.86), which are commonly used for summarization tasks. This higher score indicates that BEAR can better capture important words and phrases from the original text.

4) ROUGE-2 (R-2): With an R-2 score of 8.57, BEAR substantially surpasses the other models in capturing bigram-level co-occurrences between the generated summaries and the reference summaries. For instance, SEQ and BERTEXT which scored 5.23 and 4.48 respectively, lag behind in capturing contextual relationships that involve pairs of words. The improvement in R-2 reflects BEAR's enhanced ability to generate coherent summaries.

5) ROUGE-L (R-L): BEAR also leads in the ROUGE-L metric with a score of 19.44, demonstrating its strength in maintaining the longest common subsequence between the generated summaries and the references. This indicates that BEAR is more capable of producing summaries that retain the overall meaning and structure of the original content compared to other models like SEQ (16.15) and CPSUM (Soft) (12.71).

Overall, BEAR's results on the XSum dataset show that the model consistently outperforms the other methods across all metrics, validating its capability to generate high-quality summaries for summarization tasks.

C. Role of Clustering and Representative Sentences

Clustering organizes sentences into thematic groups based on their semantic similarity, utilizing TF-IDF and K-means clustering algorithms. This method structures the text into distinct clusters, each representing specific topics or aspects, which helps maintain context and relevance throughout the summarization process. By segmenting the text into these thematic clusters, redundancy is minimized and the focus is placed on selecting the most representative sentences. This reduces repetition and ensures a concise, relevant summary.

Once the clustering is complete, representative sentences from each cluster are selected and serve as the context for the BART model. The BART model then generates the final summary by integrating these representative sentences, this hybrid approach effectively bridges the gap between extractive and abstractive summarization techniques. It ensures that the summary not only captures key points but also maintains fluency and coherence, facilitating smooth transitions between different sections. The combination of clustering and BART helps to create a summary that is both contextually relevant and well-structured.

D. Limitations of the Present Study

While the approach aims to enhance coherence and relevance, there might still be instances where the generated summaries contain redundant or less relevant information if the clustering or similarity measures are not perfectly aligned with the summary objectives.

VI. CONCLUSION

In this paper, a novel framework for text summarization is presented that enhances the quality of generated summaries. The approach integrates a rewrite model with an innovative sentence sampling strategy, resulting in summaries that are not only more relevant but also of higher quality. A method for the rewrite model leverages an effective sentence sampling strategy, enhancing the selection of content for summarization.

TABLE V. STATE-OF-THE-ART COMPARISON FOR TEXT SUMMARIZATION

Model	CNN/DailyMail			XSUM		
	R-1	R-2	R-L	R-1	R-2	R-L
Lead 3	40.01	17.45	36.31	16.30	1.60	11.95
PacSum	40.37	17.92	36.62	-	-	-
FAR	40.42	17.95	36.67	-	-	-
SUMO	41.00	18.40	37.20	-	-	-
PGNet	39.50	17.30	36.40	-	-	-
REFRESH	41.30	18.40	37.50	-	-	-
SEQ	23.24	7.10	22.15	20.11	5.23	16.15
TED	38.73	16.84	36.15	-	-	-
BART	44.16	18.07	35.53	-	-	-
CPSUM (Soft)	-	-	-	17.22	2.17	12.71
CPSUM (Hard)	-	-	-	17.29	2.18	12.73
BERTEXT	-	-	-	22.86	4.48	17.16
BEAR (proposed model)	**48.15**	**19.23**	**37.03**	**25.00**	**8.57**	**19.44**

The framework consistently outperforms existing state-of-the-art baselines across various text summarization tasks. These results underscore the superiority of the approach in delivering more coherent and contextually relevant summaries. Future work will focus on several key areas to further enhance the summarization framework. First, the application of this approach to other languages and domains will be explored, expanding its scope and effectiveness. Additionally, methods to incorporate user feedback into the summarization process will be explored, enabling the generation of more personalized and user-centric summaries.

AUTHORS' CONTRIBUTIONS

Zahir Edrees: Conceptualization, Methodology, Software, Writing original draft, Validation, Resources, Visualization, review and editing.Yasin Ortakci: Conceptualization, Methodology, Project administration, Supervision, review and editing.

DATA AVAILABILITY STATEMENT

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

COMPETING INTERESTS

The authors have no competing interests to declare that are relevant to the content of this article.

CONFLICT OF INTERESTS

This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

ETHICS AND INFORMED CONSENT FOR DATA USED

The research does not involve human participants and/or animals.

FUNDING AND ACKNOWLEDGMENT

This research received no external funding.

REFERENCES

[1] S. Zhang, R. Yang, X. Xiao, X. Yan, and B. Tang, "Effective and efficient pagerank-based positioning for graph visualization," *Proceedings of the ACM on Management of Data*, vol. 1, pp. 1–27, 5 2023.

[2] S. Gongid, Z. Zhu, J. Qi, C. Tong, Q. Lu, and W. Wu, "Improving extractive document summarization with sentence centrality," 2022. [Online]. Available: https://doi.org/10.1371/journal.pone.0268278.g001

[3] X. Zhang, F. Wei, and M. Zhou, "Hibert: Document level pre-training of hierarchical bidirectional transformers for document summarization," *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pp. 5059–5069, 2019. [Online]. Available: https://aclanthology.org/P19-1499

[4] Y. Liu, P. Liu, D. Radev, and G. Neubig, "Brio: Bringing order to abstractive summarization," *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 2890–2903, 3 2022. [Online]. Available: https://arxiv.org/abs/2203.16804v1

[5] H. T. Kesgin and M. F. Amasyali, "Advancing nlp models with strategic text augmentation: A comprehensive study of augmentation methods and curriculum strategies," *Natural Language Processing Journal*, vol. 7, p. 100071, 6 2024.

[6] S. Pawar, H. M. Gururaj, and N. N. Chiplunar, "Text summarization using document and sentence clustering," vol. 215. Elsevier B.V., 2022, pp. 361–369.

[7] W. Liu, Y. Sun, B. Yu, H. Wang, Q. Peng, M. Hou, H. Guo, H. Wang, and C. Liu, "Automatic text summarization method based on improved textrank algorithm and k-means clustering," *Knowledge-Based Systems*, vol. 287, 3 2024.

[8] Y. Liu, "Fine-tune bert for extractive summarization." [Online]. Available: https://arxiv.org/abs/1908.08345

[9] S.-N. Vo, T.-T. Vo, and B. Le, "Interpretable extractive text summarization with meta-learning and bi-lstm: A study of meta learning and explainability techniques," *Expert Systems With Applications*, vol. 245, p. 123045, 2024. [Online]. Available: https://doi.org/10.1016/j.eswa.2023.123045

[10] M. E. Saleh, Y. M. Wazery, and A. A. Ali, "A systematic literature review of deep learning-based text summarization: Techniques, input representation, training strategies, mechanisms, datasets, evaluation, and challenges," *Expert Systems With Applications*,

vol. 252, p. 124153, 2024. [Online]. Available: https://doi.org/10.1016/j.eswa.2024.124153

[11] M. Zhong, P. Liu, Y. Chen, D. Wang, X. Qiu, and X. Huang, "Extractive summarization as text matching," pp. 6197–6208. [Online]. Available: https://github.

[12] F. Mohsen, J. Wang, and K. Al-Sabahi, "A hierarchical self-attentive neural extractive summarizer via reinforcement learning (hsasrl)," *Applied Intelligence*, vol. 50, pp. 2633–2646, 9 2020.

[13] Y. Wang, J. Zhang, Z. Yang, B. Wang, J. Jin, and Y. Liu, "Improving extractive summarization with semantic enhancement through topic-injection based bert model," *Information Processing and Management*, vol. 61, 5 2024.

[14] I. Benedetto, M. L. Quatra, L. Cagliero, L. Vassio, and M. Trevisan, "including those for text and data mining, ai training, and similar technologies. tasp: Topic-based abstractive summarization of facebook text posts," *Expert Systems With Applications*, vol. 255, p. 124567, 2024. [Online]. Available: https://doi.org/10.1016/j.eswa.2024.124567

[15] H. Shakil, A. Farooq, and J. Kalita, "To appear in: Neurocomputing," 2024. [Online]. Available: https://doi.org/10.1016/j.neucom.2024.128255.

[16] J. You, R. Ying, and J. Leskovec, "Position-aware graph neural networks." [Online]. Available: http://snap.stanford.

[17] S. A. Bahrainian, G. Zerveas, F. Crestani, and C. Eickhoff, "Cats: Customizable abstractive topic-based summarization," *ACM Transactions on Information Systems*, vol. 40, 1 2022.

[18] M.-H. Su, C.-H. Wu, and H.-T. Cheng, "A two-stage transformer-based approach for variable-length abstractive summarization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2061–2072, 2020.

[19] M. Zhang, C. Li, M. Wan, X. Zhang, and Q. Zhao, "Rouge-sem: Better evaluation of summarization using rouge combined with semantics," *Expert Systems With Applications*, vol. 237, 2024. [Online]. Available: https://doi.org/10.1016/j.eswa.2023.121364

[20] A. M. Rush, S. Chopra, and J. Weston, "A neural attention model for abstractive sentence summarization," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, L. Màrquez, C. Callison-Burch, and J. Su, Eds. Lisbon, Portugal: Association for Computational Linguistics, Sep. 2015, pp. 379–389. [Online]. Available: https://aclanthology.org/D15-1044

[21] A. See, P. J. Liu, and C. D. Manning, "Get to the point: Summarization with pointer-generator networks," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, R. Barzilay and M.-Y. Kan, Eds. Vancouver, Canada: Association for Computational Linguistics, Jul. 2017, pp. 1073–1083. [Online]. Available: https://aclanthology.org/P17-1099

[22] R. Nallapati, B. Zhou, C. N. dos santos, C. Gulcehre, and B. Xiang, "Abstractive text summarization using sequence-to-sequence rnns and beyond," 2 2016. [Online]. Available: http://arxiv.org/abs/1602.06023

[23] D. Morozovskii and S. Ramanna, "Rare words in text summarization," *Natural Language Processing Journal*, vol. 3, p. 100014, 6 2023.

[24] D. Jain, M. D. Borah, and A. Biswas, "Summarization of lengthy legal documents via abstractive dataset building: An extract-then-assign approach," *Expert Systems with Applications*, vol. 237, 3 2024.

[25] W.-W. Qiu, H.-T. Yu, C.-H. Tsai, D. Zhu, M.-H. Chen, and H. J. Kim, "Understanding the value of host-guest intimacy behind online reviews of airbnb," *International Journal of Hospitality Management*, vol. 115, p. 103599, 2023. [Online]. Available: https://doi.org/10.1016/j.ijhm.2023.103599

[26] "Artificial intelligence approach for detection and classification of depression among refugees in selected diasporic novels — enhanced reader."

[27] P. Radhakrishnan and G. SenthilKumar, "Stab: An enhanced abstractive text summarization employing stacked bi-gru with the attention cnn approach," *SN Computer Science*, vol. 5, 8 2024.

[28] M. Kirmani, G. Kour, M. Mohd, N. Sheikh, D. A. Khan, Z. Maqbool, M. A. Wani, and A. H. Wani, "Biomedical semantic text summarizer," *BMC Bioinformatics*, vol. 25, 12 2024.

[29] D. Parnes and A. Gormus, "Prescreening bank failures with k-means clustering: Pros and cons," *International Review of Financial Analysis*, vol. 93, p. 103222, 2024. [Online]. Available: https://doi.org/10.1016/j.irfa.2024.103222

[30] I. Trabelsi, R. Hérault, H. Baillet, R. Thouvarecq, L. Seifert, and G. Gasso, "Identifying patterns in trunk/head/elbow changes of riders and non-riders: A cluster analysis approach," *Computers in Biology and Medicine*, vol. 143, p. 105193, 2022. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2021.105193

[31] W. A. Prastyabudi, A. N. Alifah, and A. Nurdin, "Sciencedirect peer-review under responsibility of the scientific committee of the seventh information systems international conference segmenting the higher education market: An analysis of admissions data using k-means clustering," 2023. [Online]. Available: www.sciencedirect.comwww.sciencedirect.com

[32] "Study of anisotropy in polydispersed 2d micro and nanocomposites by elbow and k-means clustering methods — enhanced reader."

[33] N. Rylko, M. Stawiarz, P. Kurtyka, and V. Mityushev, "Study of anisotropy in polydispersed 2d micro and nanocomposites by elbow and k-means clustering methods," *Acta Materialia*, vol. 276, p. 120116, 2024.

[34] T. P. Rinjeni, A. Indriawan, and A. Rakhmawati, "Sciencedirect peer-review under responsibility of the scientific committee of the seventh information systems international conference matching scientific article titles using cosine similarity and jaccard similarity algorithm," *Procedia Computer Science*, vol. 234, pp. 553–560, 2024.

[35] T. Nakagawa, Y. Sanada, H. Waida, Y. Zhang, Y. Wada, O. Takanashi, T. Yamada, and T. Kanamori, "Denoising cosine similarity: A theory-driven approach for efficient representation learning," *Neural Networks*, vol. 169, pp. 893–6080, 2024. [Online]. Available: https://doi.org/10.1016/j.neunet.2023.10.027

[36] K. Moritz, H. Tomas, K. Kocisky, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, P. Blun-

som, and G. Deepmind, "Teaching machines to read and comprehend." [Online]. Available: http://www.github.com/deepmind/rc-data/

[37] S. Narayan, S. B. Cohen, and M. Lapata, "Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization," pp. 1797–1807.

[38] R. Barzuay, K. R. Mckeown, and M. Elhadad, "Information fusion in the context of multi-document summarization."

[39] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," *Annual Meeting of the Association for Computational Linguistics*, 2004.

[40] G. Erkan and D. R. Radev, "Lexrank: Graph-based lexical centrality as salience in text summarization," *Journal Of Artificial Intelligence Research*, vol. 22, pp. 457–479, 9 2011. [Online]. Available: http://arxiv.org/abs/1109.2128 http://dx.doi.org/10.1613/jair.1523

[41] U. Jayasankar, V. Thirumal, and D. Ponnurangam, "A survey on data compression techniques: From the perspective of data quality, coding schemes, data type and applications," *Journal of King Saud University - Computer and Information Sciences*, vol. 33, pp. 119–140, 2 2021.

[42] S. Narayan, N. Papasarantopoulos, S. B. Cohen, and M. Lapata, "Neural extractive summarization with side information." [Online]. Available: www.aaai.org

[43] I. S. Google, O. V. Google, and Q. V. L. Google, "Sequence to sequence learning with neural networks," 2014.

[44] C. Li, Q. Li, P. V. Mieghem, H. E. Stanley, and H. Wang, "Correlation between centrality metrics and their application to the opinion model," *The European Physical Journal B*, vol. 88, p. 65, 12 2015.

[45] A. Onan and H. A. Alhumyani, "Fuzzytp-bert: Enhancing extractive text summarization with fuzzy topic modeling and transformer networks," *Journal of King Saud University-Computer and Information Sciences*, vol. 36, p. 102080, 2024. [Online]. Available: https://doi.org/10.1016/j.jksuci.2024.102080

[46] G. Sharma, D. Sharma, and M. Sasikumar, "Summarizing long scientific documents through hierarchical structure extraction," *Natural Language Processing Journal*, vol. 8, p. 100080, 2024. [Online]. Available: https://doi.org/10.1016/j.nlp.2024.100080

[47] C.-Y. Lin, "ROUGE: A package for automatic evaluation of summaries," in *Text Summarization Branches Out*. Barcelona, Spain: Association for Computational Linguistics, Jul. 2004, pp. 74–81. [Online]. Available: https://aclanthology.org/W04-1013

Hiding Encrypted Images in Audios Based on Cellular Automatas and Discrete Fourier Transform

Jose Alva Cornejo, Esdras D. Vasquez, Jose Calizaya Quispe, Roxana Flores-Quispe, Yuber Velazco-Paredes
Universidad Nacional de San Agustín de Arequipa, Arequipa, Perú

***Abstract*—With the increasing need for secure long-distance communication, protecting sensitive information such as images during transmission remains a significant challenge. This paper proposes a new method for hiding encrypted images inside audio files by integrating Cellular Automata (CA) and the Discrete Fourier Transform (DFT). The primary aim is to enable secure transmission of large encrypted images without altering the audio's perceptual quality. The scheme leverages the cryptographic properties of CA to generate encrypted images, which are then embedded into inaudible frequencies of audio using DFT. Results show that this method successfully hides and recovers images of considerable size, maintaining bit-level integrity of the original images while preserving audio quality. However, the scheme lacks resilience to signal processing attacks, such as compression or filtering, the resulting size of the audio is also bigger. Despite this limitations, the method provides a competitive advantage in payload capacity and efficiency, making it suitable for applications where the transmission of large, sensitive data is necessary but not subject to aggressive signal attacks.**

***Keywords*—Cellular automaton; Fourier Transform; cryptography; synchronization; stenography; embedding**

I. Introduction

Nowadays with the development of information and communication technologies, access to information has become easier and establishing communication in a secure way has become a necessary requirement [1]. As a result, people can easily exchange information and distance is no longer a barrier to communication. However, the safety and security of long-distance communication remains an issue [2], because in many cases, the Internet is being affected by hackers.

For that reason, attempts have been made to provide a cyber security environment to protect the assets of institutions, organizations, and individuals such as encryption systems, watermarking, steganography, fingerprinting, hybrid systems [3].

It is therefore essential to investigate more secure and efficient methods for safeguarding sensitive data, such as images, during transmission over open channels.

In the case of steganographic techniques, many were proposals to provide secure data exchange through an open communication channel. These approaches are mainly hosted under three domains: In spatial domain techniques [4] [5], the data is hidden, and replacement is directly applied to the pixels of the image; Transform Domain Methods hide the messages in significant areas of the cover image to produce more efficient stego-images. It manipulates the image indirectly by various transformation techniques; the most popular of these techniques are: Discrete Cosine Transformation (DCT) and Discrete Wavelet Transformation (DWT); and the third domain considers hybrid domain techniques; which is a type of steganography where spatial and transform domains may be combined. The hybrid approaches also provide some security and capacity enhancements but still are in their beginnings and need more research. [6].

In other cases, Cellular Automata (CA) models have been used for their good cryptographic properties that provide security against attacks and better confusion and diffusion properties [7]. CA models also give a secret key for the encryption which cannot be predicted since it evolves into a chaotic and complex system starting from an initial state [8].

So the question arises How can encryption methods be integrated with a CA cellular automata to hide encrypted images in audio files to improve transmission security? With all that has been seen, it was asked, how can encryption methods be integrated with a well-encryption system like CA with steganography to improve the security of encrypted image transmission?

For that reason, this paper proposes a novel method to increase security in image transmission: A hybrid method to encrypt high quality images using CA and hide them inside audio files using the Discrete Fourier Transform (DFT) based on the methods proposed by Alvarez et al. [9] and Hwai-Tsu and Tung-Tsu [10] for improved security in hiding encrypted images. This method could be of great use in industries that require the protection of sensitive information, such as the financial, military or healthcare sectors, where data integrity and confidentiality are paramount.

The rest of the paper is organized as follows: The review of previous works is presented in Section II, the proposed scheme is described in Section III, where a description of fundamental concepts is made and then the detailed description of each step in encryption and decryption is presented, in Section IV the results are shown along with their respective analysis, the discussion is presented in Section V and it ends with the conclusion in Section VI.

II. Related Work

In recent years, significant efforts have been made to solve the problem of information security in data transmission, such as the work done by Alvarez et al. [9], who have proposed a scheme based on the bidimensional reversible CA with memory. These schemes are cryptographic procedures to share a secret among a set of participants in such a way that only some qualified subsets of these participants can recover the secret. Also, the security of the scheme is studied and it is proved that the protocol is ideal and perfect and also resists the most important statistical attacks. To validate the

protection of the original information, the number of changing pixel rate (NPCR) and the unified averaged changed intensity (UACI) randomness test were used, with scores of 99 and 33 respectively, which would indicate a high level of change in the encrypted image compared to the original. This suggests a robust encryption that is very sensitive to changes in the original image; in other words, a minimal modification in the original image will cause a noticeable difference in the encrypted image, which is positive for the security of the encryption.

Similarly, Hwai-Tsu and Tung-Tsu [10] have proposed a study to introduce an innovative phase modulation (PM) scheme based on the fast Fourier transform (FFT) that facilitates efficient and effective blind audio watermarking. The results reflected the robustness of phase modulation against a variety of common signal processing attacks, and comprehensive and rigorous tests confirmed the PM's robustness against a variety of common signal processing attacks, including resampling, requantization, and low pass filtering. But the FFT–PM was less resistant to attacks that caused severe phase perturbations.

Likewise, Eslami et al. [11] proposed a model using CA and a double authentication mechanism to propose a new threshold image sharing scheme with steganographic properties. That proposed scheme uses 2 bits in each pixel of cover images for embedding data and so a better visual quality for the produced stego images was achieved. Consequently, this study got a Peak Signal-to-Noise Ratio (PSNR) value of 48, showing that the difference between the original image and the compressed or encrypted image is minimal and that the loss of quality is practically imperceptible to the human eye.

Similarly, Hernández et al. [12] proposed a new graphic symmetrical cryptosystem in order to encrypt a colored image defined by pixels and by any number of colors. This cryptosystem is based on a reversible bidimensional CA and uses a pseudo-random bit generator where the session key is the seed used to generate the pseudorandom bit sequence. In consequence, the decrypted image is identical to the original, i.e., no loss of resolution occurs.

On the other hand, Tanwar & Bisla [13] understood that the goal of audio steganographic technique is to embed data in audio cover files that must be robust and resistant to malicious attacks. That paper presents various audio steganographic methods like LSB, echo hiding, spread spectrum etc. Also, merits and demerits of each method are described. Finally, they showed that the low bit coding (LSB) method has a low robustness, echo masking, and phase coding have a low capacity for data embedding; in relation to the spectrum, it has a higher robustness but is vulnerable with respect to the modification of the time scale.

Abdirashid, Solak & Saku [14] argued that image steganography techniques provide better data embedding capability. So, they proposed secure data hiding algorithms based on frequency domain in image steganography. The methods were evaluated according to the criteria of imperceptibility, payload capacity and robustness, where they obtained good results of PSNR of 50 dB and Structural Similarity Index Measure (SSIM), which represents a successful restoration for the same image.

Also noteworthy is the comparison of different types of encryption realized by Louis [15], which tests the DCT and DFT encryption techniques in order to compare which of the two is better where it is seen that even though the images are embedded using bytes complete directly into the Fast Fourier Transform (FFT) transformed dimension, the byte difference is distributed somewhat evenly over the entire image, with small differences of the pixel values. That method is thus harder to detect than embedding methods that directly use the spatial domain.

An interesting paper was realized by Najiya and Renjith [16] who describe a method of compressing the image using wavelet compression and converting it into a bit sequence in order to embed it in the modified cover audio using a secret key, then the audio is encoded with an error correction code to improve its robustness of this technique. Where at the receiver section, the original secret image is reconstructed successfully. The PSNR value of the image in the system is 32.84. This means that the difference between the original image and the altered image is small.

The review of related works shows that many of these methods focus on information protection using different techniques in the spatial domain, transform domain and hybrid domain, this encouraged further research and improvement in this field.

III. Proposed Method

Fig. 1 shows the representation of the proposed method hiding encrypted images by CA in audios based on DFT. And, Fig. 2 shows the process to decrypt the information.

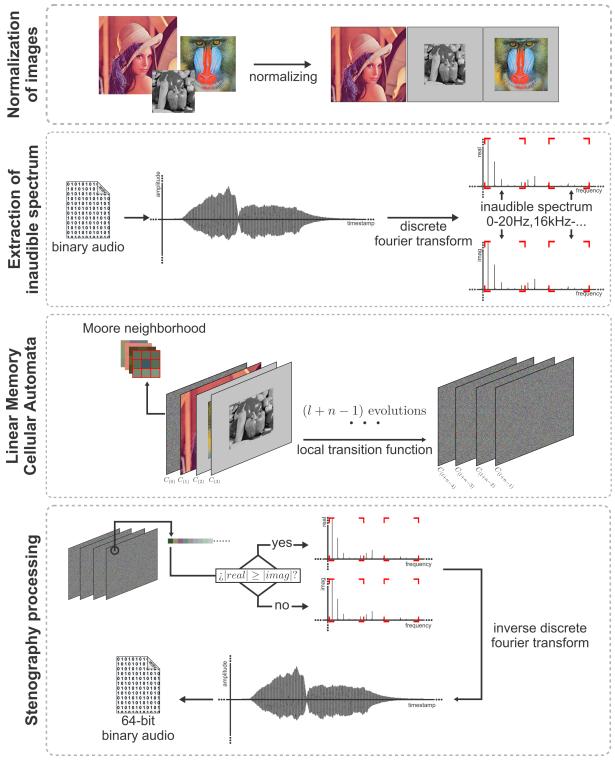


Fig. 1. Encryption of the proposed scheme.

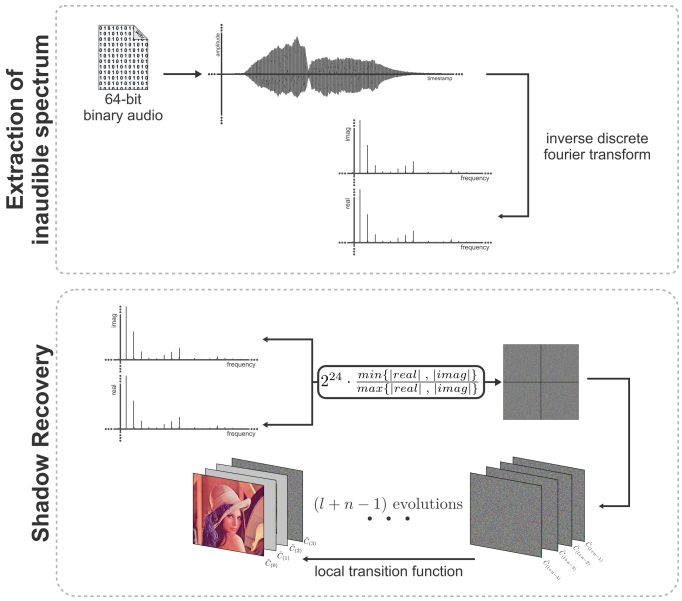


Fig. 2. Decryption of the proposed scheme.

A. Normalization of Images

This first stage is considered essential, because the operation of the CA depends on its environment, i.e. "neighborhood", based on the number of cells given by the number of rows and columns, which is simply feasible for any case. However, the proposal is to hide images by means of these evolutions, that means that the space operated by the CA must be the same in all cases of evolutions, by adding white pixels around each image if necessary.

Therefore, from a set of RGB images of different sizes, they are normalized by adding a padding of white pixels so that they have the same dimensions, Fig. 3 shows an example, not including the borders.

(a) (b) (c) (d)

Fig. 3. Normalized images.

B. Extraction of Inaudible Spectrum

Although the CA gives the necessary images to be able to reverse the process, these images could have essentially noise, which can generate suspicion, therefore, a higher stage of encryption is proposed, which means hiding it inside an audio.

The method supports the reading of different types of audio in wav format $(16-bit\ PCM,\ IEEE\ float)$, internally all processing is done in $64-bit$ float format, so any conversion from an audio with a smaller data format does not result in any alteration of the original audio [17]. Audio recorded at sampling rates greater than or equal to $44100Hz$ are used, because this is the standard used in music CDs and provides good frequency precision [18]. The output audio will be a $64-bit\ IEEE\ float\ .wav$ file.

On the other hand, of all the forms of audio steganography that exist, the decision has been to use the algorithm called "Fast Fourier Transform" or FFT, to hide data in the less perceptible frequencies: from 0 to $20Hz$ and from $16kHz$ [19].

Where, for each audio channel, it will be divided into M segments of length L each one. Using the DFT, each segment frequency is obtained using the Eq. 1.

$$X^{(m)}(k) = \sum_{n=0}^{L-1} x^{(m)}(n) \mathrm{e}^{-\mathrm{i}2\pi k \frac{n}{L}} \tag{1}$$

where $m = \{0 \dots M\}$
$k = \{0, \dots, 20, 16000, 16001, \dots \frac{L}{2}\}$
* Illustrative quantity, take into account the frequency resolution.

However, the nature of the FFT algorithm will return $N/2$ samples from $N = 2^n, n \in \mathbb{R}^+$ samples. Therefore it is set that $L = 65536 = 2^{16}$ to ensure that a wide range of frequencies is obtained. That results by the concept called frequency resolution, which is shown in Eq. 2.

$$\Delta f = \frac{f_s}{N} \tag{2}$$

f_s : Frequency sample rate
N : Number of samples

$$e.g: \ \Delta f = \frac{44100}{65536} \approx 0.67$$
$$\Delta f = \frac{44100}{32768} \approx 1.34$$
$$\Delta f = \frac{44100}{16384} \approx 2.69$$

$\therefore$ More sample gets more resolution

Then the "frequency resolution" is used to put the encrypted data into an array, in this way it can approximate a 1:1 relationship between the indices of an array and frequencies it want to access.

Additionally, a function is used to access a frequency by indices of an array, which is shown in Eq. 3:

$$freqToIndex(freq) = \lfloor \frac{freq}{2^{16}} \rfloor, \quad freq \geq 0 \tag{3}$$

C. Linear Memory Cellular Automata

CA is a computational model that simulates dynamic systems and processes within a concrete space composed of cells. These cells have different states that vary in relation to time due to predefined rules. Each cell of a CA is composed of a dynamic state in relation to its neighboring cells by means of specific rules, which define the state configuration, simulating the evolution of a system over multiple continuous time steps [9].

According to [9], the following elements define a CA:

- State It is the set S of possible values that a cell will have. This paper makes use of RGB images leads to a $S = \mathbb{Z}_{2^{24}}$, it is the minimum value containing all possible values of an RGB pixel.
- Local transition function Function that gives a new state to the cell (i, j) from its neighborhood V_{ij}, using a number ω of 9 bits defined as a rule. It is defined in the following way:
$$s_{ij}^{(t+1)} = F(V_{ij}^{(t)}, ..., V_{ij}^{(t-n)})$$
- Neighborhood It is defined as the set of neighbors of a cell. Therefore, in this paper the Moore neighborhood is used.

Instead of working on a one-dimensional line, two-dimensional CA are organized in a grid or matrix of cells, where each cell can have a particular state, and offers a way to model and understand complexity starting from simple local rules. Their versatility and ability to represent a wide variety of events make them a valuable tool [20].

The change of the evolutions depends on the states previous to these since are based on the $k - th$ order, and each of them defines a new change of states of the CA based on the k previous evolutions. For example, if k takes the value of 3, then the new state in $(t + 1)$, depends on the states of (t), $(t - 1)$ and $(t - 2)$.

Now, the security of the evolutions to be performed in the Linear Memory Cellular Automata (LMCA) depends on an image composed of random pixels, indicating dimensions and other characteristics, see Fig. 4. This will be considered as configuration 0, C_0 while the images to be encrypted will be the images shown in Fig. 3: C_1 C_2 C_3 C_4 respectively. Based on the fact that this has $n = 4$ initial images, along with an initial configuration, the order of the cellular automaton would be 5; which implies that the 5 previous ones will always be used to generate the next one (times), and so on until the number of evolutions [9][21].

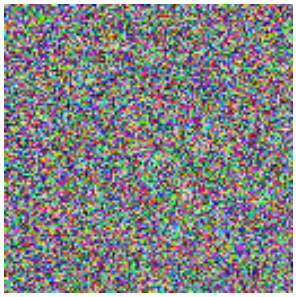


Fig. 4. Configuration C_0.

For each time it have, w is a random number. According to the model of an LMCA, a random number l is necessary, which determines the number of evolutions defined as $\#evo = n + l - 1$, for the present investigation the value of l will be equal to 10 ($l = 10$), the results of which are shown in Fig. 5.

At the moment of making the evolutions, it can be observed how images of what is apparently noise (shadows) are created,

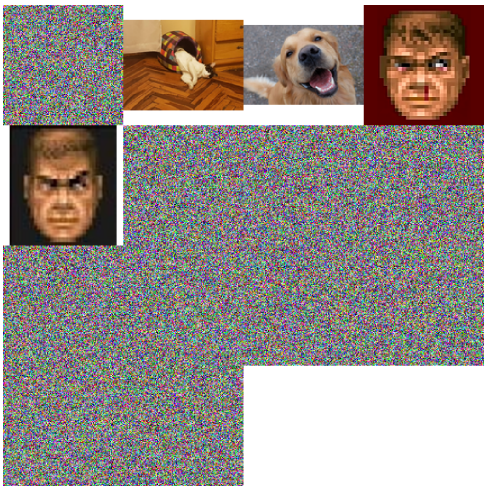


Fig. 5. Evolutions with a value of l=10.

and according to [11], the correlation between shadows is very small, which guarantees that it is not possible to extract relevant information from the original images by having only a fraction of the shadows. The LMCA has a set of times, because it used is a k to k symmetric scheme, the keys for the recovery will be composed by the same number of initial configurations as it had. In the example, the last four noise images would be presented together with the one preceding them. The latter would serve as a public key, while the other four are secret.

D. Stenography Process

It is widely acknowledged that the human auditory system exhibits relatively low sensitivity to variations in phase [22]. For this reason, the key to imperceptible audio primarily involves manipulating the FFT phase. Since each FFT coefficient comprises a real and an imaginary component, this scheme for numerical embedding involves the manipulation of the ratio between the magnitudes of these two components. The component with the largest magnitude was identified as the baseline unit. Consequently, the extent of the other component is modulated based on the intended numeric value, such as a pixel value extracted from a color image. Now, the Eq. 4 is derived from the equation proposed by [10]. This equation allows encrypting an entire RGB color.

$$\begin{aligned}
&\text{if } |\text{Re}\{X^{(m)}(k)\}| \geq |\text{Im}\{X^{(m)}(k)\}| \\
&\quad \hat{X}^{(m)}(k) = \text{Re}\{X^{(m)}(k)\} \\
&\qquad + i{\cdot}sgn(\text{Im}\{X^{(m)}(k)\}) \cdot |\text{Re}\{X^{(m)}(k)\}| \cdot \frac{v(\Delta + k)}{2^{24}} \\
&\text{else} \\
&\quad \hat{X}^{(m)}(k) = sgn(\text{Re}\{X^{(m)}(k)\}) \cdot |\text{Im}\{X^{(m)}(k)\}| \cdot \frac{v(\Delta + k)}{2^{24}} \\
&\qquad + i \cdot \text{Im}\{X^{(m)}(k)\}
\end{aligned} \tag{4}$$

where $v(x)$ refers to the value of the ID array that contains the pixels of an image.
$\Delta = m(L - 16020)$ storage capacity in a block m.

Once the values of each pixel of an image have been positioned inside the inaudible frequencies of an audio, the Inverse Fourier Transform (see Eq. 5) is operated to obtain an audio that will be saved in a 64-bit .wav file in order to safeguard the precision of the mathematical operations and the values of this new audio.

$$\hat{x}^{(m)}(k) = \frac{1}{L}\sum_{n=0}^{L-1} \hat{X}^{(m)}(n)\mathrm{e}^{\mathrm{i}2\pi k\frac{n}{L}} \tag{5}$$

This whole process is shown in Algorithm 1.

Algorithm 1 Encryption

Require: D is a vector of data to encrypt
Require: A is a $m \times n$ matrix where

$$m = \begin{cases} 1 & \text{if audio channel is mono} \\ 2 & \text{if audio channel is stereo} \end{cases}$$

and $n = Length\ of\ audio$

Ensure: fft(...) returns an array in CCs format
  $res \leftarrow A\ m \times n\ matrix\ with\ encrypted\ data$
  $samples \leftarrow 2^{16}$
  $encrypted \leftarrow 0$ ▷ Progress
  for all $signal$ **in** A **do**
    $segment \leftarrow 0$
    while $segment \leq |signal|$ **and** $encrypted \neq |D|$ **do**
      $out \leftarrow$ fft$(signal[segment{:}])$ ▷ out is a vector of $\frac{samples+1}{2}D$
      $out \leftarrow \frac{out}{samples}$ ▷ Normalize
      for $i \leftarrow 1$ **to** $|out| - 1$ **and** $encrypted \neq |D|$ **do**
        if $i = freqToIndex(20.0)$ **then** ▷ See equation 3
          $i \leftarrow freqToIndex(16000.0)$
          continue
        end if
        $x \leftarrow out[i]$
        if $|real(x)| \geq |imag(x)|$ **then** ▷ See equation 4
          $imag(x) \leftarrow sgn(imag(x)) * |real(x)| * \frac{D[encrypted]}{2^{24}}$
        else
          $real(x) \leftarrow sgn(real(x)) * |imag(x)| * \frac{D[encrypted]}{2^{24}}$
        end if
        $encrypted \leftarrow encrypted + 1$
      end for
      $out\_inv \leftarrow$ ifft(out)
      $copy\ out\_inv\ into\ res$
    end while
    $copy\ signal[segment{:}]\ to\ res[signal]$ ▷ Remainder
  end for
  return res

E. Shadow Recovery

Similar to the previous step, the recovery is based on separating the audio stego into segments of 2^{15} audio samples, where in each one the FFT is applied to recuperate the values of the frequencies where the data is saved.

$\tilde{X}^{(m)}(k)$ is define as the sequence of values obtained by using the FFT in a block; the same variables defined in the previous step are shown by Eq. 6.

$$v(\Delta + k) = 2^{24} \cdot \frac{min\{|\mathrm{Re}\{\tilde{X}^{(m)}(k)\}|\ ,\ |\mathrm{Im}\{\tilde{X}^{(m)}(k)\}|\}}{max\{|\mathrm{Re}\{\tilde{X}^{(m)}(k)\}|\ ,\ |\mathrm{Im}\{\tilde{X}^{(m)}(k)\}|\}} \tag{6}$$

After getting the pixel array, they are saved for later use in the CA.

Continuing, the images have been recovered from the audio files and a process identical to the first evolution is developed where the public key will be the configuration 0, and the following shadows are added in the order they were generated, by effecting the number of evolutions with the same value l, the original images are obtained (with their respective padding. See Fig. 6 and 7.

This whole process is shown in Algorithm 2.

Algorithm 2 Decryption

Require: A is a $m \times n$ matrix where

$$m = \begin{cases} 1 & \text{if audio channel is mono} \\ 2 & \text{if audio channel is stereo} \end{cases}$$

and $n = Length\ of\ audio$

Require: $size$: The desired size of the decrypted data
Ensure: fft(...) returns an array in CCs format
  $res \leftarrow$ Array of decrypted data
  $decrypted \leftarrow 0$ ▷ Progress
  for all $signal$ **in** A **do**
    $segment \leftarrow 0$
    while $segment + 2^{16} < |signal|$ **and** $decrypted \neq size$ **do**
      $out \leftarrow$ fft$(signal[segment{:}])$
      $out \leftarrow out/samples$ ▷ Normalize
      for $i \leftarrow 1$ **to** $|out| - 1$ **do**
        if $decrypted = size$ **then**
        end if
        if $i = freqToIndex(20.0) + 1$ **then**
          $i \leftarrow freqToIndex(16000.0)$ **continue**
        end if
        $x \leftarrow out[i]$
        $v \leftarrow \lfloor \frac{min\{|real(x)|,|imag(x)|\}}{max\{|real(x)|,|imag(x)|\}} * 2^{24} \rfloor$ ▷ See Eq. 6
        $res.append(v)$ ▷ Store decrypted value
        $decrypted \leftarrow decrypted + 1$
      end for
    end while
  end for
  return res

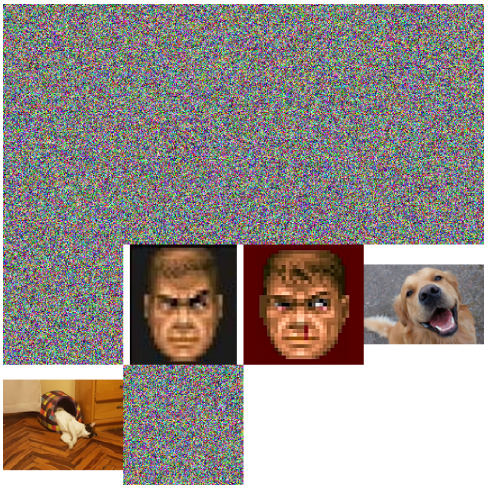


Fig. 6. Inverse evolution.

Fig. 7. Images recovered.

IV. Results

A. Random Noise NIST Test Results

To ensure that the secret images are indeed noise, the NIST statistical test suite is used, it consists 15 different tests that measure the randomness of a sequence of bits. For each test a p-value is calculated, which varies from 0 to 1, where 0 indicates that the bits are non random and 1 perfect randomness [23].

To test the first step of the scheme, a large image(4000x3000 pixels) is used to generate the appropriate noise images to hide. For this example, nine evolutions are created; and then, for each images, a binary file composed entirely of the pixel RGB bytes is made.

Each of the nine files passed to the NIST suite contain 288 million bits and are of size 36MB. The results in Table I give them an average p-value of a little under 0.5, which shows good randomness. There were only a couple of failed tests in the earlier evolutions, however, it can be concluded that the secret images to be hidden in an audio file will be very close to random noise.

TABLE I. Results of Randomness Test

Evolution	Average p-value	Passrate
1	0.467008	99.47%
2	0.495432	100.00%
3	0.481725	98.94%
4	0.471555	99.47%
5	0.478481	100.00%
6	0.502455	100.00%
7	0.457972	100.00%
8	0.474596	100.00%
9	0.499604	100.00%

B. Size and Capacity of Audio Files

Because the process of hiding images in audios using the FFT internally uses $64 - bit$ floating numbers, it cannot save the frequency values in audio formats such as $16 - bit$ fixed-point and $32 - bit$ floating-point without losing data in the truncation process, which forces to save $64 - bit$ IEEE wav audios. See Table II to view the comparison of the created audios.

TABLE II. Size Analysis of Created Audios

	Audio1	**Audio2**	**Song1**	**Song2**
Duration	4s	7s	6:23 min	2:49 min
Orig Size	812.6 KiB	1.5 MB	64.5 MB	28.5 MB
64bit Size	3.2 MB	5.8 MB	257.9 MB	114.1 MB
Increase	293.798%	286.667%	299.845%	300.351%
Max Capacity (NxN)	256	330	2153	1434

It can be observed that the created audios are about four times larger, but it allows them to save an RGB image of considerable size, or in the case of smaller images, a set of these.

C. Spectrograms

By performing a spectrogram analysis of both the original audio (see Fig. 8) and the one containing their payload (see Fig. 9), it can be observed how it proceeds according to the proposed method, where only the less audible frequencies have been covered by what amounts to noise.

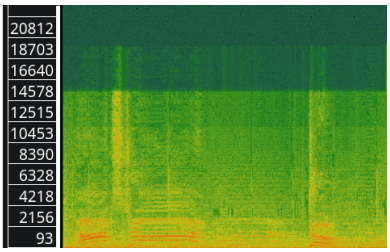


Fig. 8. Spectrogram of original audio.

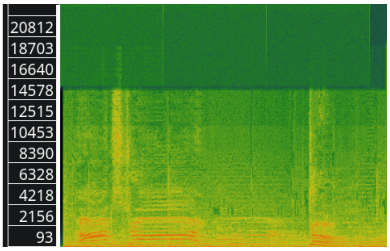


Fig. 9. Spectrogram of stego-audio.

In a second sample only a 128×128 pixel image has been inserted, it can be observed that only a section of the high frequencies is modified, this due to the fragmentation of the audio for the realization of the process, see Fig. 10 and 11.

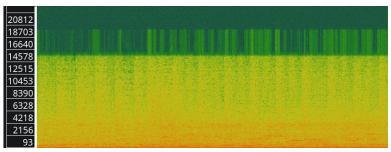


Fig. 10. Spectrogram of second original audio.

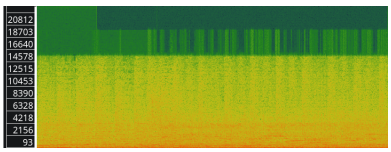


Fig. 11. Spectrogram of second stego-audio.

D. Measuring Similarity to Original Audio

1) Dynamic Time Warping: As a way of measuring how similar the payloaded audio is to the original one, the algorithm

Dynamic Time Warping (DTW) has been used. DTW is a really powerful tool with uses way beyond just measuring differences in audio files, such as speech and sign language recognition, computer vision, animation, data mining, music and signal processing [24]. DTW compares two signals that may or may not be of the same length to find an optimal alignment with minimal cost, to do this a cost matrix is found, by comparing every value of both signals, were the higher the difference between values, the higher the cost. After obtaining the matrix, the optimal warping path is the one with the lowest accumulated cost from traversing from bottom left to top right [25]. Because the signals to be compared are practically of the same length, the alignment path will always be a straight line from bottom to top.

To find a measurement related to the similarity of both audio signals, two comparisons have been decided upon. The first comparison uses DTW on the original audio and the stego audio. The second comparison is performed between the original signal and a noisy signal, which is generated by adding almost imperceptible random noise to the original. This approach provides a baseline for assessing how different the stego signal is.

Fig. 12 and 13 show the cost matrix of both comparisons, where orange tones show a greater cost. The shortest path is always the direct one, as a better way of showing the differences of the comparison, Fig. 14 and 15 shows a slice of the cost matrices, but plotted as a 3d bar chart, the first comparison presents lower values overall and by calculating the distance cost of the optimal warping path. The values obtained are 1528.1893 for the stego audio and 82826.3947 for the noisy signal. This indicates that the stego audio signal is easier to adapt and correct, and thus very similar to the original audio.

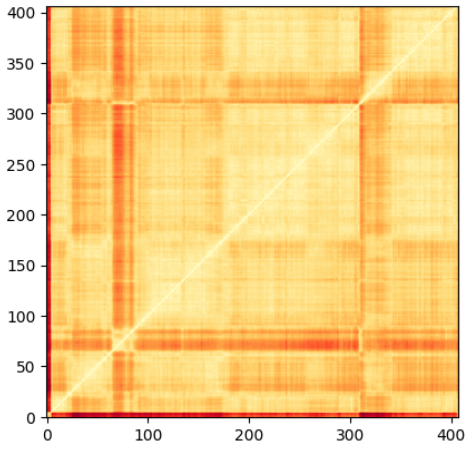


Fig. 12. DTW cost matrix of stego-audio comparison.

2) Error Testing: An evaluation will be made with the Mean Square Error (MSE) and Signal-to-Noise Ratio (SNR) coefficients to see the efficiency of the proposed method, by the use of the Fourier transform to audio.

In the MSE, the error signal $e_i = x_i - y_i$ represents the difference between the original and distorted signals [26], which will be evaluated with respect to audio quality, after performing data hiding within the audio during the transform. See Eq. 7.

$$MSE = \frac{\sum(O - E)^2}{N} \tag{7}$$

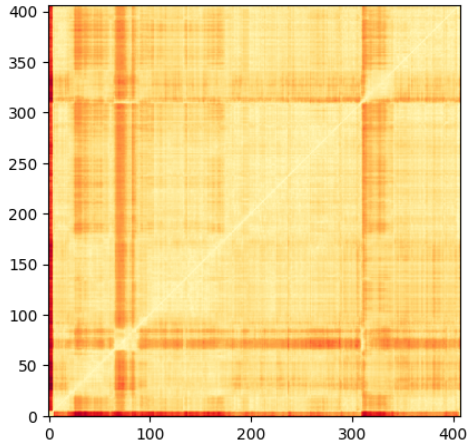


Fig. 13. DTW cost matrix of noisy audio comparison.

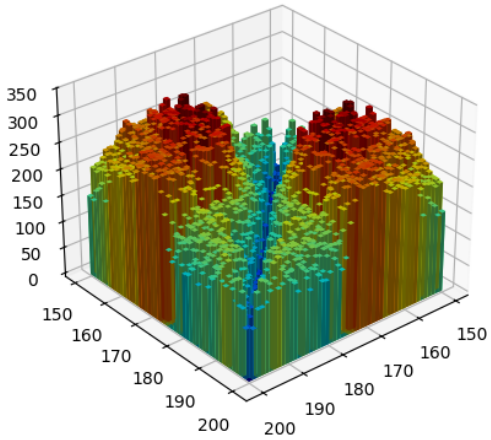


Fig. 14. 3D bar chart of DTW cost matrix of stego-audio comparison.

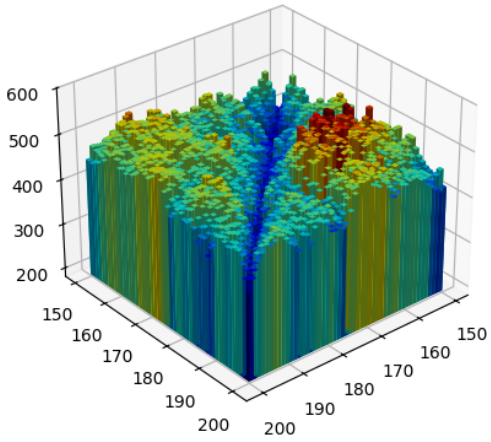


Fig. 15. 3D bar chart of DTW cost matrix of noisy-audio comparison.

On the other hand, SNR reflects the signal-to-noise relation of an audio, usually written as S/N, is an estimate of the robustness of the source signal in relation to the possible noise (unwanted signal) [27], see Eq. 8

$$SNR = 10 * \log_{10}(\frac{O^2}{MSE}) \tag{8}$$

After several comparisons of encrypted and decrypted audios, The resulting values are shown in the Table III.

TABLE III. ANALYSIS OF ORIGINAL AND ENCRYPTED AUDIO

Sample Indices	Original (O)	Encrypted (E)	(O^2)	(O - E)^2
0	0	-0.0048	0.0048000	2.28 $\times 10^{-5}$
6500	0.2461	0.2510	0.060500	2.44 $\times 10^{-5}$
13000	0.0080	0.0113	0.0000634	1.13 $\times 10^{-5}$
19500	-0.1737	-0.1757	0.0302000	4.11 $\times 10^{-6}$
26000	0.1949	0.2014	0.0380000	4.26 $\times 10^{-5}$
32500	0.1617	0.1589	0.0261000	7.75 $\times 10^{-6}$
39000	0.0972	0.0927	0.0094300	2.09 $\times 10^{-5}$
45500	0.0008	-0.0098	0.0000007	1.12 $\times 10^{-4}$
52000	-0.2294	-0.2190	0.0526000	1.08 $\times 10^{-4}$
58500	0.0896	0.0811	0.0080300	7.30 $\times 10^{-5}$
65000	-0.0730	-0.0728	0.0053300	7.22 $\times 10^{-8}$
71500	0.1217	0.1218	0.0148000	2.02 $\times 10^{-8}$
78000	-0.1572	-0.1565	0.0247000	5.69 $\times 10^{-7}$
84500	-0.0857	-0.0888	0.0073400	9.72 $\times 10^{-6}$
91000	0.0306	0.0295	0.0009340	1.08 $\times 10^{-6}$
97500	-0.0554	-0.0560	0.0030700	3.33 $\times 10^{-7}$
104000	-0.0363	-0.0365	0.0013200	3.67 $\times 10^{-8}$
110500	-0.0547	-0.0512	0.0029900	1.23 $\times 10^{-5}$
117000	-0.0070	-0.0077	0.0000497	4.86 $\times 10^{-7}$
123500	0.0403	0.0422	0.0016300	3.59 $\times 10^{-6}$
130000	0.0410	0.0404	0.0016800	4.13 $\times 10^{-7}$
136500	-0.0800	-0.0824	0.0064100	5.68 $\times 10^{-6}$
143000	0.0247	0.0249	0.0006100	3.68 $\times 10^{-8}$
149500	0.0616	0.0613	0.0038000	8.06 $\times 10^{-8}$
156000	0.0523	0.0520	0.0027300	1.02 $\times 10^{-7}$
162500	-0.2395	-0.2397	0.0573000	3.63 $\times 10^{-8}$
169000	0.1144	0.1134	0.0131000	1.03 $\times 10^{-6}$
175500	0.0418	0.0435	0.0017500	2.90 $\times 10^{-6}$
182000	0.0649	0.0637	0.0042200	1.63 $\times 10^{-6}$
188500	0.0071	0.0074	0.0000501	8.14 $\times 10^{-8}$
195000	-0.0119	-0.0114	0.0001410	2.42 $\times 10^{-7}$
201500	0.0072	0.0072	0.0000523	4.9 $\times 10^{-14}$
208000	0.0054	0.0054	0.0000295	2.8 $\times 10^{-14}$
Total			**0.37908**	**0.00047**

The difference of the amplitude of the original audio minus the encrypted audio is solved for each row of the table and the result elevated to 2.

All the results in column 5 of Table III are summed, and the value of the variable N is equal to the number of samples that exist in the table, in this case 33. Therefore using the Eq. 7

$$MSE = \frac{\sum(O-E)^2}{N}$$
$$MSE = \frac{0.00047}{33}$$
$$MSE = 0.0000142$$

This MSE result affirms that there is a minimum change between the original audio and the encrypted audio because the audio is close to 0, and as explained above, the closer the MSE is to zero, the smaller the average difference will be; therefore, the better the quality of the audio reproduction or processing, so it is reaffirmed that the audio quality is maintained after applying the FFT algorithm to the audio.

The summation of O^2 is in column 4 of Table III, and the value of the MSE is the result obtained previously, therefore using the Eq. 8:

$$SNR = 10 * \log_{10}(\frac{0.37908}{0.0000142})$$
$$SNR = 10 * \log_{10}(26695.77)$$
$$SNR = 10 * 4.43$$
$$SNR = 44.26dB$$

This result of 44.26 dB, reaffirms that the quality of the signal (sound) is preserved, after doing the Transform algorithm, so it deduced that the sound emitted by the audio remains clear and neat after encryption, and that any evident change between the values of the original audio with the encrypted audio is not perceptible or is minimally perceptible.

E. Recovered Images Similarity

By way of a more detailed analysis of the decrypted images in order to ensure pixel equality, MSE and SSIM measurements are used in the comparison of the original image (with some padding) and the one resulting from the encryption and decryption process. See the Table IV.

TABLE IV. IMAGES SIMILARITY MEASUREMENTS

	MSE	SSIM
Cat	0.00	1.00
Dog	0.00	1.00
Doom1	0.00	1.00
Doom2	0.00	1.00

MSE is used, because of its ample use and its simplicity to compare two images [28], every single image recovered by the scheme got a perfect score of 0, which indicates that the squared sum of the errors always equal 0, therefore, every single pixel value of both the recovered images and the original ones are the same.

The use of SSIM helps to evaluate the similarity between original and distorted images after applications in the evaluation of image quality, image recovery, image encryption and data hiding [29]. In effect, a value of 1 was obtained as result, this is evidence that the images still maintain the exact same image quality before and after encryption.

V. DISCUSSION

The use of Cellular Automata in the scheme ensures the security of the sensitive images to be hidden. These images are converted to effectively noise, any attacker that manages to extract the data hidden in the very low and high frequencies would find nothing but useless data.

In this study, an SNR of 44.26 has been obtained using the DFT together with CA. This result is notably superior to that obtained by another research on the embedding of color images in audio signals using residual networks [10]. In that work, the SNR ranged from 18.75 to 25.004, depending on the embedding scheme and the range of FFT indices employed.

An SNR of 44.26 indicates that the audio signal has a significantly higher SNR than the watermarking schemes employed in [10]. This suggests that the method using DFT and CA can insert information into the audio signal with less degradation of the perceived quality, which is crucial in applications where maintaining the fidelity of the original signal is a priority.

The presented scheme shows good payload capacity for the resulting file sizes, a 4 seconds audio file is capable of hiding an image of at most 256x256 dimensions, with a final size of 3.2MB. When comparing it to the scheme proposed by [10], a similar size audio file (4.23MB) of 24.15 seconds of duration, is able to hide an image of 64x64 dimensions. This shows a considerable increase in capacity per MB, albeit with a very reduced resistance to signal processing attacks and compression techniques because of the need to preserve every bit of the secret CA images to be able to decrypt the original images. Because of this, the proposed scheme should probably be used on relatively small audio files, so that the increase in size is not as dramatic and any action that affects the audio file data, such as applying mp3 compression, should be considered destructive to the encrypted information.

VI. Conclusion

Based on the challenge of secure and private information sharing, a new encryption method using CA and Fourier Transform based steganography has been proposed in this research. This method demonstrates the ability to preserve the full integrity of the encrypted images while maintaining an extremely similar quality to the original audio, where the hidden data is no different to noise, supported SNR, MSE and a magnitudes lower optimal warping path length compared to an audio containing an almost imperceptible noise.

Because of the combination of the produced audios file format and the range of frequencies that are embedded with data, the proposed scheme has good payload capacity, albeit with the limitation of the resulting file size and the vulnerability to small changes in the audio file.

Therefore, as possible improvements for future work, different approaches could be explored, such as implementing support for lower and higher color depth images at $24 - bits$, optimizing the Fast Fourier Transform process to use less precision, which would allow us to hide encrypted images in smaller floating point audio formats such as 32 and 24 bits, reducing the audio file size at the expense of data storage capacity.

References

[1] S. Karakus and E. Avci, "A new image steganography method with optimum pixel similarity for data hiding in medical images," *Medical Hypotheses*, vol. 139, p. 109691, 2020.

[2] N. Hamid, A. Yahya, R. B. Ahmad, D. Najim, and L. Kanaan, "Steganography in image files: A survey," *Australian Journal of Basic and Applied Sciences*, vol. 7, pp. 35–55, 2013.

[3] O. Tayan, "Concepts and tools for protecting sensitive data in the it industry: A review of trends, challenges and mechanisms for data-protection," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 2, 2017.

[4] M. Hussain, A. W. A. Wahab, Y. I. B. Idris, A. T. Ho, and K.-H. Jung, "Image steganography in spatial domain: A survey," *Signal Processing: Image Communication*, vol. 65, pp. 46–66, 2018.

[5] S. Solak and U. Altinişik, "Image steganography-based gui design to hide agricultural data," *Gazi University Journal of Science*, vol. 34, no. 3, pp. 748–763, sep 2021.

[6] J. W. M. C. P. W. B. C. 2, "A survey on digital image steganography," *Journal of Information Hiding and Privacy Protection*, vol. 1, no. 2, pp. 87–93, 2019. [Online]. Available: http://www.techscience.com/jihpp/v1n2/29000

[7] F. E. Ziani, A. Sadak, C. Hanin, B. Echandouri, and F. Omary, "Ca-pcs: A cellular automata based partition ciphering system," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, 2020. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2020.0110376

[8] S. BOUCHKAREN and S. LAZAAR, "A fast cryptosystem using reversible cellular automata," *International Journal of Advanced Computer Science and Applications*, vol. 5, no. 5, 2014. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2014.050531

[9] G. Alvarez, L. Hernández Encinas, and A. Martín del Rey, "A multisecret sharing scheme for color images based on cellular automata," *Information Sciences*, vol. 178, no. 22, pp. 4382–4395, 2008.

[10] H.-T. Hu and T.-T. Lee, "Hiding full-color images into audio with visual enhancement via residual networks," *Cryptography*, vol. 7, no. 4, 2023.

[11] Z. Eslami, S. Razzaghi, and J. Zarepour, "Secret image sharing based on cellular automata and steganography," *Pattern Recognition*, vol. 43, pp. 397–404, 01 2010.

[12] G. A. M. L. H. E. A. H. E. A. M. del Rey and G. R. Sanchez, "Graphic cryptography with pseudorandom bit generators and cellular automata," *Digital.CSIC*, 2002.

[13] R. Tanwar and M. Bisla, "Audio steganography," pp. 322–325, 2014.

[14] A. Mohamed Abdirashid, S. Solak, and A. K. Sahu, "Data hiding based on frequency domain image steganography," no. 42, p. 71–76, 2022.

[15] F. I. Louis, "Image steganography in the frequency domain," 2023.

[16] N. T. E1 and R. V. Ravi, "An effective technique for hiding image in audio," *International Journal of Science and Research*, vol. 4, 2015.

[17] I. D. A. Focus and T. W. Groups, "Recommended practices for enhancing digital audio compatibility in multimedia systems," *DATWG Recommendation*, 1992.

[18] D. J. Katz and R. Gentile, "Chapter 5 - basics of embedded audio processing," in *Embedded Media Processing*, ser. Embedded Technology, D. J. Katz and R. Gentile, Eds. Burlington: Newnes, 2006, pp. 149–187. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780750679121500081

[19] G. Plenge, H. Jakubowski, and P. Schöne, "Which bandwidth is necessary for optimal sound transmission?" *Journal of the Audio Engineering Society*, vol. 28, no. 3, pp. 114–119, 1980.

[20] P. Lazzari and N. Seriani, "Two-dimensional cellular automata—deterministic models of growth," *Chaos, Solitons & Fractals*, vol. 185, p. 114997, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960077924005496

[21] D. Anani and K. M. Faraoun, "Designing robust lmca-based threshold secret sharing scheme for digital images using multiple configurations assignment," *Journal of Communications Software and Systems (JCOMSS)*, vol. 11, no. 2, JUNE 2015.

[22] F. E. Toole, "Sound reproduction loudspeakers and rooms," *ELSEVIER Ltd*, 2008.

[23] L. Bassham, A. Rukhin, J. Soto, J. Nechvatal, M. Smid, S. Leigh, M. Levenson, M. Vangel, N. Heckert, and D. Banks, "A statistical test suite for random and pseudorandom number generators for cryptographic applications," 2010-09-16 2010. [Online]. Available: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=906762

[24] P. Senin, "Dynamic time warping algorithm review," *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, vol. 855, no. 1-23, p. 40, 2008.

[25] *Dynamic Time Warping*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 69–84. [Online]. Available: https://doi.org/10.1007/978-3-540-74048-3_4

[26] W. Zhou and A. Bovick, "Mean squared error, love it or leave it?" *IEEE SIGNAL PROCESSING MAGAZINE*, 2009.

[27] M. R. Orora Tasnim, Selim Hossain, "Audio steganography with intensified security and hiding capacity," *Eur. Chem Bull Section A-Research paper*, 2023.

[28] U. Sara, M. Akter, and M. S. Uddin, "Image quality assessment through fsim, ssim, mse and psnr—a comparative study," *Journal of Computer and Communications*, vol. 07, no. 03, p. 8–18, 2019.

[29] H. R. S. Zhou Wang, Alan Conrad Bovik and E. P. Simoncelli, "Image quality assesment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

DBPF: An Efficient Dynamic Block Propagation Framework for Blockchain Networks

Osama Farouk[1], Mahmoud Bakrey[2], Mohamed Abdallah[3]
Department of Information System-Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt, 41522[1,2]
School of Computing and Data Science, BADYA University, Cairo, Egypt[3]

***Abstract*—Scalability poses a significant challenge in blockchain networks, particularly in optimizing the propagation time of new blocks. This paper introduces an approach, termed "DBPF" - Dynamic Block Propagation Framework for Blockchain Networks, aimed at addressing this challenge. The approach focuses on optimizing neighbor selection during block propagation to mitigate redundancy and enhance network efficiency. By employing informed neighbor selection and leveraging the Brotli lossless compression algorithm to reduce block size, the objective is to optimize network bandwidth and minimize transmission time. The DBPF framework calculates the Minimum Spanning Tree (MST) to ensure efficient communication paths between nodes, while the Brotli compression algorithm reduces the block size to optimize network bandwidth. The core objective of DBPF is to streamline the propagation process by selecting optimal neighbors and eliminating unnecessary data redundancy. Through experimentation and simulation of the block propagation process using(DBPF), we demonstrate a significant reduction in the propagation time of new blocks compared to traditional methods. Comparisons against approaches such as selecting neighbors with the least Round-Trip Time *RTT*, random neighbor selection, and the DONS approach reveal a notable decrease in propagation time up to more than (45%) compared to them based on network type and number of nodes. The effectiveness of (DBPF) in boosting blockchain network efficiency and decreasing propagation time is emphasized by the experimental findings. Additionally, various compression algorithms such as zstandard and zlib were tested during the research. Nevertheless, the results suggest that Brotli produced the most positive outcomes. Through the integration of optimized neighbor selection and effective data compression, DBPF presents a hopeful resolution to the scalability issues confronting blockchain networks. These results showcase the capability of (DBPF) to notably enhance network performance, leading the path toward smoother and more efficient blockchain operations**

Keywords*—Blockchain; scalability; minimum spanning tree; compression; broadcasting; optimized neighbor selection; network bandwidth; transmission time optimization

I. Introduction

Blockchain (BC) technology has revolutionized the way data and financial transactions are shared among network participants BC was initially introduced in 2008, credited to Satoshi Nakamoto [1]. At its core, a blockchain is a decentralized and distributed ledger that securely records transactions across a network of nodes. The value added by blockchain technology lies in its ability to provide transparency, immutability, and trust. By eliminating the need for intermediaries and central authorities, blockchain enables peer-to-peer transactions and data sharing, fostering a new era of decentralized applications and services [2], [3]. This improves its reliability and efficiency compared to traditional data storage systems. The networks within (BC) can securely manage information and protect it from tampering, even in the presence of some malicious nodes, These features are highly valuable and find applications not only in cryptocurrencies but also across a wide range of fields [4]. Therefore, (BC) has many applications in emerging fields such as the internet of things [5], [6], 5G [7] [8], [9], and artificial intelligence, [10], [11].

However, despite its numerous advantages, blockchain faces scalability challenges that hinder its widespread adoption [12]. One of the key scalability issues is the low throughput of blockchain networks, primarily caused by the high propagation time required to disseminate new blocks to all nodes in the network. This delay in block propagation can lead to inconsistencies in the blockchain, impacting transaction finality and network performance. It is evident that Bitcoin's transaction processing capability is limited, allowing for only seven Transactions Per Second (TPS). This stands in stark contrast to widely adopted mainstream payment platforms like PayPal, which achieves a transfer rate of 500 TPS, and Visa, surpassing 4000 TPS [13]. Similarly, **Ethereum**, another prominent cryptocurrency, can handle approximately 15 TPS. Clearly, both **Bitcoin** and **Ethereum** are inadequate when it comes to meeting the demands of large-scale trading scenarios.

The high propagation time in blockchain networks can be attributed to several factors. Firstly, traditional methods used for path selection and neighbor discovery in blockchain networks often result in suboptimal routing paths, leading to longer propagation times and message redundancy. Random Neighbor Selection (RNS) is one such method that may contribute to inefficient block dissemination by selecting neighbors without considering the network topology or optimal paths, where shared data propagates through random paths [14], leading to an inefficient data propagation scheme. This inefficiency arises from the probability of redundancy in the exchanged messages between network nodes. This redundancy arises from cycling in randomly chosen data paths, leading to longer delivery times and reduced consistency. Despite this, most blockchain (BC) systems support Random Neighbor Selection (RNS). Several methods have been proposed to enhance the Neighbor Selection (NS) process locally, addressing the issue of dynamicity. For example, Bi et al. [15] introduced an NS protocol based on network latency, where nodes measure the Round Trip Time (RTT) to their neighboring nodes and prioritize those with the lowest RTT for the NS process. However, none of these methods has presented an optimal NS strategy.

Additionally, the size of blocks and network bandwidth

limitations also play a significant role in increasing propagation time. As block sizes grow, the time required to propagate blocks across the network increases proportionally [16]. Moreover, some blockchain protocols overlook the optimization of network bandwidth usage, further exacerbating propagation delays.

In response to these challenges, this paper proposes a novel solution to enhance blockchain scalability by addressing the inefficiencies in neighbor selection and block size management. Our approach focuses on optimizing neighbor selection for each node in the network (ONS) by calculating the Minimum Spanning Tree (MST) to establish efficient communication paths. Furthermore, we aim to optimize network bandwidth by reducing block size and improving propagation speed by implementing the (Brotli compression) method, a lossless data compression technique that minimizes redundant data within blocks and transactions. By combining optimized neighbor selection and block size reduction through compression, our proposed solution (**DBPF**) seeks to mitigate the propagation time challenges in blockchain networks, ultimately enhancing network scalability and performance.

The contributions of this paper are summarized as follows:

1) The Dynamic Block Propagation Framework (DBPF) addresses blockchain scalability by optimizing neighbor selection through a Minimum Spanning Tree (MST), reducing redundant data propagation, and improving communication efficiency between nodes.
2) DBPF incorporates Brotli lossless compression to reduce the size of the propagated blocks, improving network bandwidth usage and speeding up block dissemination.
3) By leveraging MST-based Optimal Neighbor Selection, DBPF selects neighbors that enable faster and more efficient block propagation, reducing the overall time needed to disseminate blocks across the network.
4) The framework significantly reduces the total propagation time for data exchanged between nodes, ensuring faster and more efficient block distribution.
5) DBPF combines optimized neighbor selection with effective data compression, providing a scalable and efficient solution for large blockchain systems to increase the throughput of the network.

The remaining sections of this paper are structured as follows: Section II analyzes relevant literature, Section III provides the proposed system model for DBPF, Section IV provides a detailed explanation of the proposed DBPF, Section V presents the evaluation of DBPF, and finally, Section VI summarizes the most significant findings and conclusions.

II. Related Work

This section presents several modern network layer scalability solutions, which primarily aim to enhance either the gossip algorithm or reduce block data size. Research studies targeting improvements in the gossip algorithm focus on reducing duplicate data and increasing block propagation speed [17]. These proposed solutions aim to minimize duplication caused by the gossip protocol decrease block propagation time through enhanced gossip mechanisms and reduce the size of the block to accelerate block propagation through network bandwidth. Below are some recent works that represent such solutions.

The authors of "PiChu: Accelerating Block Broadcasting in Blockchain Networks with Pipelining and Chunking" [18] introduce an innovative method to enhance the efficiency and scalability of blockchain networks. PiChu leverages pipelining and chunking techniques to expedite block propagation, significantly reducing the time required to broadcast blocks across the network. The approach involves verifying the block header before dividing the block into smaller segments, which are then processed and forwarded in parallel. This continuous data flow minimizes delays associated with waiting for entire blocks to be verified. The chunking process splits each block into smaller, self-contained chunks that include complete transactions, allowing for incremental verification and faster dissemination. PiChu is particularly effective in scenarios with high transaction volumes and large block sizes, making blockchain networks more robust and scalable. However, PiChu's effectiveness is influenced by the network topology and the connectivity between nodes. The requirement for modifications to existing consensus protocols may also limit its immediate applicability in some blockchain systems. Additionally, the benefits of pipelining and chunking may vary depending on the specific characteristics of the network, and The need for nodes to verify and forward chunks can introduce additional processing overhead.

Baniata and Anaqreh [19] introduced the Dynamic Optimized Neighbor Selection Algorithm (DONS) to enhance P2P network management within blockchain networks. In this approach, a leader peer is selected to manage the network and construct its topology using neighbor lists from regular peers. The leader uses a Minimum Spanning Tree (MST) to identify optimal neighbors, thereby minimizing propagation delay and enhancing transaction throughput. However, when the leader changes, the network topology must be reconstructed, necessitating the collection of neighbor lists anew. As the number of peers increases, the time required to compute the MST also increases, leading to inefficient bandwidth utilization. Furthermore, the unavailability of the leader poses risks of topology loss and incurs overheads associated with leader reselection.

Zhang et al. [20] investigate the dynamics of block propagation in blockchain-based vehicular networks (VANETs) and the impact of node mobility on blockchain consensus mechanisms. The authors propose an analytical model to derive a closed-form expression for single-block propagation time and analyze multi-block competitive propagation to address blockchain forking issues. The model accounts for the dynamic connectivity of moving nodes, introducing opportunistic communication to blockchain consensus. Their findings reveal that higher mobility and more moving vehicles expedite block propagation, thereby reducing consensus time. The study also shows that distinct propagation capabilities of moving nodes help mitigate blockchain forking. However, the model has limitations, such as its reliance on closed-form expressions and the assumption of a single-chain structure. The approach may face limitations in scenarios with highly dynamic topologies and varying node densities, which could affect its overall performance.

The paper "FastChain: Scaling Blockchain System with

Informed Neighbor Selection" by Ke Wang and Hyong S. Kim [21] introduces the FastChain protocol to enhance blockchain scalability by optimizing block propagation times. FastChain operates by leveraging nodes with higher bandwidth capacity to distribute blocks throughout the network. Nodes with limited bandwidth prioritize connections with high-bandwidth nodes and disconnect from those with bandwidth below a specific threshold. This mechanism comprises two essential stages: the bandwidth monitoring phase and the neighbor update phase. During the bandwidth monitoring phase, each node maintains a table containing the recent bandwidth information of its neighboring nodes. In the neighbor update phase, nodes periodically update their connections, disconnecting from those with low bandwidth. This informed neighbor selection policy significantly reduces block propagation time, thereby increasing the effective block rate and improving throughput by 20% to 40%. Despite its advantages, the FastChain protocol introduces additional processing overhead due to the continuous monitoring and updating of neighbor connections. Moreover, reliance on high-bandwidth nodes could lead to centralization risks within the network.

The authors in [22], [23] proposed a score-based neighbor selection protocol for constructing a blockchain network. This protocol assigns higher scores to peers with lower propagation delays compared to those with higher propagation delays. Subsequently, peers with the highest scores are chosen as neighbors. Each miner node evaluates its neighboring nodes based on the time difference between when the block was created and when it was received by the recipient node. After successfully receiving ten blocks, a node updates its list of neighbors. During this update, the node randomly selects new neighbors, including only those with high scores. Neighbor nodes that exhibit faster block transfer rates compared to others are assigned higher scores, indicating superior network communication capabilities. Thus, miners prefer these high-scoring neighbors. This method leads to excessive dependence on the nodes that have the shortest total propagation time, which can reduce node performance.

Wang introduces the Txilm protocol in [24] to tackle the challenge of large data transmissions in blockchain networks. This is achieved through the application of lossy block compression alongside salted short hashing. The protocol operates by pre-sorting transactions based on their identifiers (TXIDs) or other criteria and hashing them using a short hash function combined with a cryptographic salt. This process significantly reduces the data size, resulting in substantial bandwidth savings. The compressed transaction list is broadcast instead of the original transactions, achieving up to 100x bandwidth efficiency. In cases of hash collisions, a second-stage resolution involving Merkle tree recomputation ensures data integrity. While the protocol effectively reduces data transmission sizes and enhances network scalability, it introduces additional computational overhead for collision resolution and depends on the consistent ordering of transactions and mempool size across nodes.

III. The Proposed System Model

To address the scalability and efficiency challenges in blockchain networks, we propose the Dynamic Block Propagation Framework (DBPF). This framework integrates MST-based optimal neighbor selection with Brotli compression for block data, aiming to minimize propagation time and enhance network throughput.

Let $G = (V, E)$ represent the blockchain network, where V is the set of nodes and E is the set of edges. Each edge $(i, j) \in E$ is associated with a weight w_{ij}, indicating the latency or communication cost between nodes i and j. The MST $T = (V, E_T)$ is a subset of E that spans all nodes V with the minimum total edge weight. The MST is computed using Kruskal's algorithm, formulated as:

$$\min \sum_{(i,j)\in E_T} w_{ij}$$

The MST provides the optimal neighbor selection (ONS_i) for each node i.

Let B denote the block containing T transactions. The block size is given by:

$$\|B\| = \sum_{i=1}^{T} \|t_i\|$$

Applying Brotli compression reduces the block size to $C(B)$, with the compression ratio $R = \frac{\|B\|}{C(B)}$. The propagation time P_t for a block B is determined by the compressed block size $C(B)$ and the network bandwidth W:

$$P_t = \frac{C(B)}{W}$$

For the entire network, the total propagation delay Δt considering the MST is given by:

$$\Delta t = \sum_{(i,j)\in E_T} \frac{C(B)}{W_{ij}}$$

where W_{ij} is the bandwidth between nodes i and j.

The overall objective of DBPF is to minimize the total propagation time by integrating MST-based optimal neighbor selection and Brotli compression:

$$\min_{T,C(B)} \Delta t = \min_{T,C(B)} \sum_{(i,j)\in E_T} \frac{C(B)}{W_{ij}}$$

This model demonstrates how DBPF optimizes both the network topology and data propagation processes, thereby enhancing blockchain network efficiency and scalability.

IV. The Proposed Efficient Dynamic Block Propagation Framework (DBPF)

This section introduces the explanation and design of an Efficient Dynamic Block Propagation Framework for Blockchain Networks (DBPF), which incorporates a combination of an efficient algorithm in neighbor selection by obtaining optimal neighbor for each node in the network by constructing a

minimum spanning tree (MST) and block size management by using **Brotli** compression algorithm to optimize network bandwidth. The primary objective of the DBPF protocol is to reduce message propagation time, optimize network bandwidth utilization, elevate overall network performance by leveraging compression techniques on block messages, and optimize blockchain network construction to be a solution for blockchain scalability problems.

To provide a clear understanding of the proposed solution, a diagram in Fig. 1 is presented. This diagram highlights the main components and mechanisms of the solution (DBPF), focusing on optimized neighbor selection and block size reduction through compression.

A. Phase-1: Leader Selection

In the context of the "**DBPF**", the leader Selection process adopts a random selection approach to designate a node as the leader within the blockchain network. This phase encompasses two primary tasks: Network leader selection and leader announcement. The DBPF requires a global view of the BC network. All nodes in BC have equal privileges in the public and permissionless BC network. However, the proposed **DBPF** selects one of these nodes to be the leader node (**LN**) to perform **MST** calculations for all other nodes. **LN** have advantages compared to other nodes in the same (BC) network, allowing it a global view or metadata information of the entire network. Additionally, **LN** collects information from the other nodes within the same network and uses it to generate the **MST** for the entire Network. Thus, each node in the network can select its optimal neighbors (ON) from the generated MST to exchange new blocks or transactions.

1) Step 1: Random selection: The **DBPF** protocol incorporates a random selection process for choosing a **LN**, which offers several notable advantages. Firstly, this approach ensures fairness by giving every node an equal opportunity to be selected as a leader, thereby eliminating potential biases. Secondly, the random selection method promotes unbiased network operations and decentralization, as it prevents the influence of network hierarchies or power dynamics in the leader selection process. This aligns with the core concept of decentralization within the BC ecosystem. Thirdly, in dynamic networks where nodes frequently join or leave, the ability to swiftly transition leadership is crucial. Random selection facilitates rapid leadership transitions without the delays associated with traditional election processes. If a **LN** departs or becomes less effective, a new leader can be randomly chosen, ensuring minimal disruption to network operations. This adaptability is paramount for maintaining the seamless functioning of the network in the face of changing circumstances.

In addition to the previously mentioned advantages, the random selection process in the **DBPF** protocol also enhances the security of the network. By randomly selecting a leader, the protocol introduces an element of unpredictability in the leadership position. Potential attackers or malicious actors are unable to predict or target specific nodes that might have elevated privileges as leaders. This randomness adds an extra layer of security to the network, as it prevents adversaries from exploiting any inherent vulnerabilities associated with predictable leader selection methods.

2) Step 2: Assessment of the Randomly Selected Leader: After a node is randomly selected as a potential leader, it undergoes a critical evaluation process to confirm its suitability. This assessment is vital to ensure that the node meets key criteria necessary for effective leadership. If the selected node meets these criteria, it is confirmed as the leader. If it does not, the random selection process is repeated. This step is designed to minimize the likelihood of choosing an unsuitable node, particularly aiming to avoid light nodes in favor of full nodes for leadership roles.

The first essential criterion is computational strength. The prospective leader node must possess substantial computational power to handle the responsibilities associated with leadership. This includes processing large volumes of data, making rapid and accurate decisions, and managing complex algorithms essential for optimizing the network. High computational capacity ensures that the node can perform its tasks efficiently without bottlenecks. Robust connectivity is another crucial requirement for the leader node. The selected leader must exhibit strong connectivity within the network. This robust connectivity is vital for ensuring efficient communication between the leader and other nodes, facilitating the swift and reliable dissemination of directives and information. Well-connected nodes enhance the overall responsiveness and coordination within the network. Finally, stability and reliability are essential characteristics of a leader node. The node must demonstrate a history of stable and reliable operation. High uptime and consistent availability are crucial, as they indicate the node's capability to maintain its leadership role without causing disruptions. A stable leader node ensures continuity and reliability in the network's functioning. To minimize the frequency of reselection, there is a preference for choosing from full nodes rather than light nodes. Full nodes, typically being more robust in terms of resources and connectivity, are more likely to meet the leadership criteria on the first selection. This approach streamlines the selection process, making it more efficient and reducing the time and resources spent in finding a suitable leader.

3) Step 3: Leadership Redetermination: In cases where the initially selected node does not meet the required criteria, the protocol stipulates a reselection. This process is promptly initiated to ensure that network leadership is established without undue delay.

Finally, The leader announcement proposed by the DBPF framework can be described as follows: Following the leader selection process, the DBPF notifies all nodes in the network about the new leader by sending announcement messages to all of them. Additionally, it informs the new leader of their responsibility for creating the MST for the network and broadcasting it to all nodes within the BC network. this phase algorithm is presented in Algorithm 1.

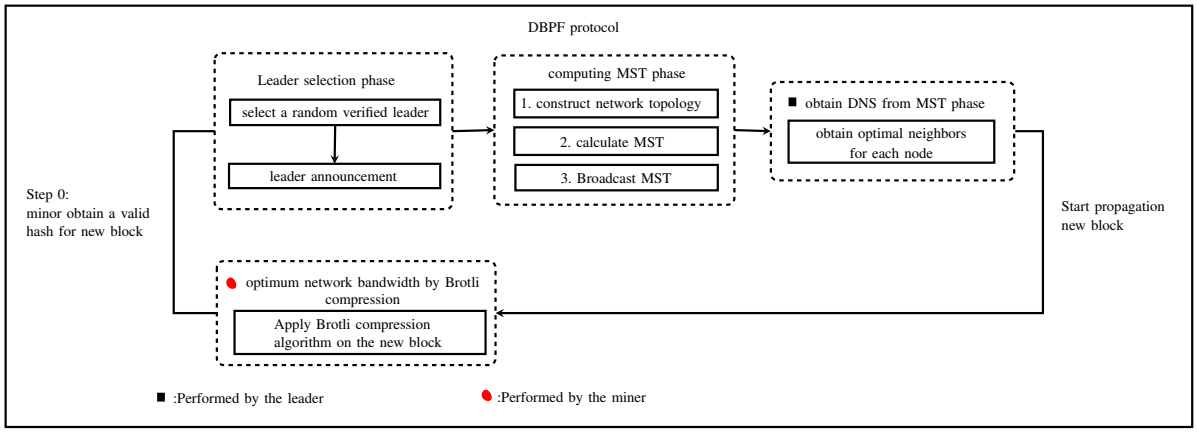


Fig. 1. The main steps involved in the proposed DBPF framework.

Algorithm 1 Leader Selection in DBPF Protocol

Require: Blockchain Network $\mathcal{N}$
Ensure: Selected Leader Node L
Begin
RandomlySelectNode:
  $n \leftarrow$ select a node randomly from $\mathcal{N}$
if $\neg$ isFullNode(n) **then**
    goto step 2 (RandomlySelectNode)
end if
AssessCandidate:
if hasComputationalStrength(n) $\wedge$ hasRobustConnectivity(n) $\wedge$ isStableAndReliable(n) **then**
    $L \leftarrow n$
else
    goto step 2 (RandomlySelectNode)
end if
End

B. Phase-2: Calculate and Dissemination of the Minimum Spanning Tree (MST)

In DBPF protocol, the subsequent phase following the random selection of the leader (L) involves a critical process that lays the foundation for network optimization. Upon phase 1, Each node and miner within the network acknowledges and recognizes the appointed leader (L). Consequently, Phase-2 is triggered and is performed by the leader (*L*) to compute MST for BC network (G) by using Kruskal's Algorithm as follows:

1) Step-1: Collection of Local information (LIs): *L* commences the process by collecting Local information (*LIs*) from all participating nodes (n) in the network (G). *LIs* consist of anonymized data regarding the network's topology, crucial for understanding the current network structure without compromising the privacy of individual nodes.

2) Step-2: Constructing Network Topology(NT): *L* employs the (*LIs*) stored locally to formulate an anonymized representation of the global network topology *NT*, Presented as an Adjacency List [25]. The Adjacency List is a fundamental data structure used to represent graph structures efficiently, particularly suitable for sparse graphs like BC networks, where the number of edges is significantly less than the number of possible edges. In the context of the Dynamic Block Propagation Framework (DBPF), the Adjacency Listis employed to construct the network topology (NT) based on the Local Information (LIs) collected from all participating nodes in the blockchain network. An Adjacency List represents a weighted graph $G = (V, E)$ where V is the set of vertices (nodes) and E is the set of weighted edges (connections with associated weights between nodes). Each vertex in the graph maintains a list of its adjacent vertices along with the weights of the connecting edges. This list captures all nodes directly connected to it by an edge and the corresponding weights, thus providing a compact and efficient way to store and manipulate the network topology.

Formally, for a weighted graph G with vertices $V = \{v_1, v_2, \ldots, v_n\}$ and weighted edges E, the Adjacency List is an array of lists, where the i-th element of the array corresponds to vertex v_i and contains a list of pairs (v_j, w_{ij}) such that there exists a weighted edge $(v_i, v_j) \in E$ with weight w_{ij}.

Example:

Consider a blockchain network with five nodes: $A, B, C, D,$ and E. The weighted connections (edges) between these nodes are as follows: $A - B(2), A - C(3), B - D(4), C - D(1), D - E(5)$. The Adjacency List representation of this weighted network would be:

- $A : \{(B, 2), (C, 3)\}$
- $B : \{(A, 2), (D, 4)\}$
- $C : \{(A, 3), (D, 1)\}$
- $D : \{(B, 4), (C, 1), (E, 5)\}$
- $E : \{(D, 5)\}$

This structure efficiently represents the network, allowing for quick look-up and traversal of adjacent nodes along with the weights of the connecting edges. Fig. 2 illustrates this weighted blockchain network graph.

The Adjacency List offers several advantages in the context of the DBPF. It provides an efficient way to store and access

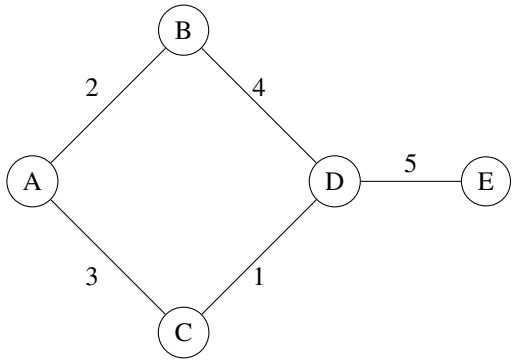


Fig. 2. Example of a weighted blockchain network graph.

the network structure, which is crucial for large blockchain networks with potentially thousands of nodes. As the blockchain network grows, the Adjacency List can handle the increasing number of nodes and edges without significant performance degradation, demonstrating its scalability. Furthermore, the list can be easily updated as new nodes join or leave the network which makes **DBPF** dynamic, ensuring that the representation remains accurate and up-to-date, highlighting its flexibility.

3) Step-3: Computation of the (MST) Using Kruskal's Algorithm:: Following the previous step where the leader node (L) utilized Local Information (LIs) to construct the network topology (NT), the next phase in the Dynamic Block Propagation Framework (DBPF) involves computing the (MST) for the blockchain network (G) using Kruskal's algorithm [26] as shown in Algorithm 2. Let $G = (V, E)$ be the graph where V is the set of vertices and E is the set of edges with weights $w : E \to \mathbb{R}$. The goal is to find a subset $T \subseteq E$ that forms a tree covering all vertices V with the minimum total edge weight:

$$\min \sum_{e \in T} w(e)$$

subject to T forming a connected acyclic subgraph of G.

Kruskal's algorithm sorts the edges E in non-decreasing order by weight and iteratively adds the shortest edge to T, provided it does not form a cycle, using the union-find data structure to manage disjoint sets. This structured approach ensures the MST is computed efficiently, forming a key component of our DBPF for optimizing blockchain network performance. The primary goal is to minimize the total edge weight while preserving network connectivity and avoiding cycles. Kruskal's algorithm processes each edge independently, which is advantageous when the number of edges $|E|$ is much smaller than the number of possible edges $\left(\frac{V(V-1)}{2}\right)$. This reduces the overall computational burden. The computed MST serves as the backbone for block propagation, significantly reducing propagation time and enhancing overall network efficiency. The time complexity of Kruskal's algorithm is $O(E \log E + E \log V)$, making it efficient for the large-scale and sparse nature of blockchain networks.

Algorithm 2 Compute MST using Kruskal's Algorithm

```
Input: Network topology NT = (V, E)
Output: Minimum Spanning Tree (MST)
procedure COMPUTE_MST(NT)
    Convert NT to a dictionary of dictionaries
    Initialize T ← ∅   ▷ Initialize the set for MST edges
    Create a union-find data structure for V
    Extract nodes and edges from NT
    Sort the edges E in non-decreasing order of their weights w(e)
    for each edge e = (u, v, w) in sorted E do
        if FIND(parent, u) ≠ FIND(parent, v) then
            T ← T ∪ {e}   ▷ Add edge to MST
            UNION(parent, rank, u, v)
        end if
    end for
    return T   ▷ The set of edges in the MST
end procedure
```

Example

Consider a blockchain network with five nodes: A, B, C, D, E. The weighted connections (edges) between these nodes are as follows: $A - B(2)$, $A - C(3)$, $B - D(4)$, $C - D(1)$, $D - E(5)$, as shown in Fig. 3.

Edge	Weight
(A, B)	2
(A, C)	3
(B, D)	4
(C, D)	1
(D, E)	5

The steps to compute the MST are as follows:

1) **Initialization**: Nodes: {A, B, C, D, E}; Edges: {(A, B, 2), (A, C, 3), (B, D, 4), (C, D, 1), (D, E, 5)}
2) **Sorting**: Sorted Edges: {(C, D, 1), (A, B, 2), (A, C, 3), (B, D, 4), (D, E, 5)}
3) **Edge Selection**:
   - Add (C, D, 1): No cycle
   - Add (A, B, 2): No cycle
   - Add (A, C, 3): No cycle
   - Skip (B, D, 4): Forms a cycle
   - Add (D, E, 5): No cycle
4) **Result**: MST: {(C, D, 1), (A, B, 2), (A, C, 3), (D, E, 5)} as shown in Fig. 4.

4) Step-4: Optimal Neighbor Selection (ONS):: Following the computation of the Minimum Spanning Tree (MST) by the leader node (L) in the previous step, the subsequent step involves disseminating the MST to all nodes within the blockchain network. This step aims to enable each node to derive its Optimal Neighbor Selection (ONS) from the received MST. Upon receiving the MST, each node runs Algorithm 3 to identify its optimal neighbors. This algorithm examines the MST and selects the most efficient connections based on the weights of the edges. These optimal neighbors are then used by the nodes to efficiently transmit new blocks or transactions across the blockchain network.

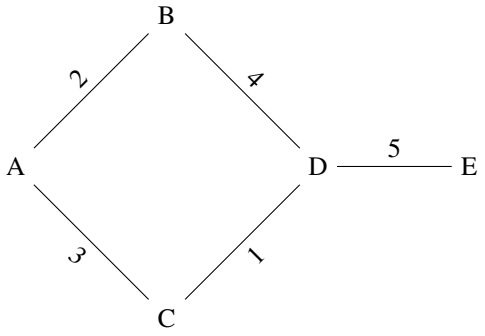


Fig. 3. Example of a weighted blockchain network graph.

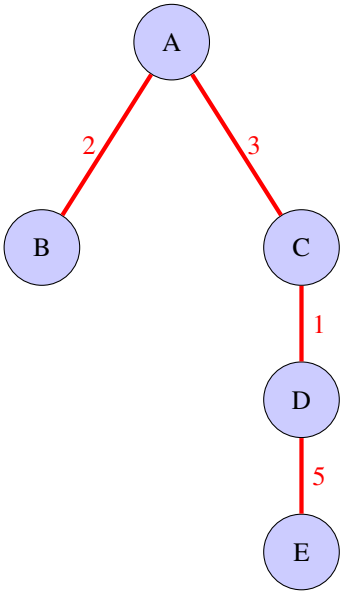


Fig. 4. Minimum spanning tree for the blockchain network.

This process ensures that data sharing is optimized, leveraging the MST to enhance overall network efficiency and performance. By utilizing the MST, the nodes can avoid redundant connections and minimize latency, which is crucial for maintaining the speed and reliability of blockchain operations. The MST-based ONS provides each node with a clear, efficient pathway for data transmission, contributing to the overall scalability and robustness of the blockchain network.

Algorithm 3 Find Optimal Neighbor Selection (ONS)

Require: MST $G = (V, E)$, node identifier node_id
Ensure: Optimal Neighbor Selection (ONS)
ONS $\leftarrow \{\}$ ▷ Initialize the set for ONS
for each $u \in G$ **do**
  if $u ==$ node_id **then**
    for each neighbor $v_j \in G[u]$ **do**
      ONS$[v_j] \leftarrow G[u][v_j]['\text{weight}']$
    end for
    break
  end if
end for
return ONS

C. Phase-3: Optimize Network Bandwidth by Brotli Compression

In this phase of the Dynamic Block Propagation Framework (DBPF), the focus shifts to optimizing network bandwidth and enhancing block propagation speed through the use of the Brotli lossless compression Algorithm [27]. This phase operates in parallel with the leader node's computation of the Minimum Spanning Tree (MST), enabling each node to transmit the compressed block efficiently to its Optimal Neighbors (ON). By doing so, the DBPF aims to reduce the propagation time of newly mined blocks across the network and increase the throughput of the public blockchain which contributes to solving the scalability problem.

The primary objective of compressing the block is to reduce its size ($\|B\|$), thereby minimizing the transmission time (P_t) across the network and optimizing the utilization of network bandwidth (W). This reduction in size aids in the faster propagation of the block to all nodes, addressing scalability issues and enhancing network efficiency.

We selected the Brotli compression algorithm for several reasons. Brotli, developed by Google, offers a high compression ratio and rapid decompression speed, making it particularly well-suited for The nature of blockchain data, which includes repetitive elements such as transaction inputs, outputs, and digital signatures, benefits greatly from Brotli's efficient compression techniques. A blockchain block is composed of two main parts: the header and the body. The header includes critical information necessary for the block's integrity and validation, such as the previous block hash, Merkle root, timestamp, difficulty target, and nonce. The body contains the transaction data, including inputs, outputs, and digital signatures. The block header's attributes are crucial for the validation cycle and maintaining the blockchain's integrity; thus, they are not compressed. Compressing these attributes would complicate validation and introduce security risks. Therefore, only the transaction data in the block body is compressed, preserving the essential information in the block header for seamless and secure validation. This compression process involves identifying redundant or less critical data within the block, applying the Brotli algorithm to reduce the block size, and then propagating the compressed block to neighboring nodes

The compression process involves several key steps. Initially, the miner identifies redundant or less critical data within the block, such as transaction inputs, outputs, digital signatures, and intermediate branches of the Merkle tree. Although the Merkle root must remain uncompressed for validation purposes, other components are suitable for compression. Let B represent a block containing T transactions, with the block size $\|B\|$ given by:

$$\|B\| = \sum_{i=1}^{T} \|t_i\|$$

Applying the Brotli compression algorithm, the block size is reduced to $C(B)$:

$$C(B) = \text{Brotli}(\|B\|)$$

The compression ratio R is defined as:

$$R = \frac{\|B\|}{C(B)}$$

This ratio reflects the effectiveness of the compression while ensuring that critical elements required for block validation, remain intact. Compressed transaction data is decompressed by nodes upon receipt, ensuring the block's integrity and validity are preserved.

When a miner successfully mines a new block B, it is compressed using the Brotli algorithm. This step occurs concurrently with the MST computation by the leader node. The miner then propagates the compressed block $C(B)$ through the optimized network to its neighboring nodes (N) which is calculated in phase 2. Each neighboring node decompresses $C(B)$ upon receipt to validate and add the block to its local blockchain. The propagation time (P_t) as a function of the compressed block size and network bandwidth is given by:

$$P_t = \frac{C(B)}{W}$$

By minimizing $C(B)$ through Brotli compression, the propagation time (P_t) is optimized, leading to enhanced network efficiency. Considering the network as a graph $G = (V, E)$, where V represents nodes and E represents communication links, the objective is to minimize the total propagation delay (Δt) across the network. This delay is influenced by the sum of individual propagation delays between nodes:

$$\Delta t = \sum_{i,j \in E} \frac{C(B)}{W_{ij}}$$

Here, W_{ij} represents the bandwidth between nodes i and j. By effectively compressing block data, DBPF optimizes the total propagation delay, improving the overall efficiency and scalability of the blockchain network.

By combining these two phases—optimized neighbor selection and enhanced network bandwidth—into a single framework called the Dynamic Block Propagation Framework (DBPF), we can significantly increase the throughput of blockchain networks. The DBPF leverages an MST-based optimal neighbor selection to establish efficient communication paths, alongside Brotli compression to reduce block sizes and transmission times. This dual approach not only minimizes propagation time but also enhances overall network efficiency. As we will demonstrate in the next evaluation section, DBPF outperforms many other methods in terms of scalability and performance in public blockchain networks.

V. Experiments and Results

This section provides a comprehensive overview of the experiments and assessments carried out to evaluate the Dynamic Block Propagation Framework (DBPF). It includes a discussion of the network datasets utilized, the performance metrics employed, and the detailed experiments executed. The network data was generated through a simulator that creates random network topologies using the Barabási-Albert(BA) model [28]. This model is well-suited for simulating real-world networks such as the Internet, social networks, and the World Wide Web.

To validate our analysis in real-world scenarios, we varied the network size and the average number of neighbors per miner. The simulation involves generating a random blockchain network where a miner node is randomly chosen as the source node for a data block. This source node disseminates the block to its neighboring nodes, which in turn propagate it to their neighbors, creating a cascading dissemination effect. The simulation ends once the block has reached all nodes in the network.

The experiments were run on a DELL PC equipped with an Intel(R) Core(TM) i5-10210U CPU (8 cores, 2.11 GHz), 8 GB DDR4-SDRAM, and a 500 GB SSD, running Windows 10 OS. We evaluated the experimental results using the Total Propagation Time (TP) metric in microseconds (μs), which measures the time it takes for a block sent from a randomly selected miner node to reach all nodes in the network.

Our experiments aim to demonstrate that the DBPF solution significantly enhances network performance by reducing the propagation time of new blocks, thereby improving the scalability and efficiency of blockchain networks.

The experiments are categorized as follows:

Experiment 1:

This experiment compares the performance of DBPF in terms of Total Propagation Time (TP) and the number of exchanged blocks (NB), against other algorithms such as Dynamic Optimized Neighbor Selection (DONS), Round-Trip Time (RTT), and Random Neighbor Selection (RNS). DONS is designed to optimize the propagation delay by selecting the most efficient paths based on delay metrics. RTT selects neighbors based on the minimum round-trip time, aiming to reduce latency in block propagation. RNS, on the other hand, randomly selects neighbors without considering network topology or path efficiency.The experiments were conducted using a network simulator based on the Barabási-Albert (BA) model to generate random network topologies. Two configurations were tested: networks with an average of 7 neighbors per node and networks with an average of 15 neighbors per node. The network bandwidth was assumed to be 10,000 bits per second. In each experiment, a randomly selected miner propagated a new block, and the propagation time and number of exchanged blocks were recorded.

The results demonstrate that DBPF significantly outperforms DONS, RTT, and RNS in both configurations. For networks with seven neighbors per node, DBPF achieved a total propagation time of 105.89 ms for 50 nodes, compared to 268.89 ms for DONS, 2124.58 ms for RTT, and 7489.9 ms for RNS. This trend continued as the network size increased to 100 and 450 nodes, with DBPF maintaining the lowest propagation times. Similarly, for networks with 15 neighbors per node, DBPF's propagation time remained consistently lower than the other algorithms. For example, with 450 nodes, DBPF achieved 156.95 ms compared to 840.50 ms for DONS, 19918.79 ms

TABLE I. PERFORMANCE OF DBPF AGAINST DONS, RTT, AND RNS

Nodes	Avg. Neighbors	Total Propagation Time (ms)			
		DBPF	DONS	RTT	RNS
50	7	105.89	268.89	2124.58	7489.9
100	7	263.37	721.72	5790.9	21089.09
450	7	326.42	1089.54	34805.4	162987.06
50	15	109.97	301.84	2152.55	14527.94
100	15	94.97	361.81	3271.19	21095.96
450	15	156.95	840.50	19918.79	125417.10

for RTT, and 125417.10 ms for RNS. The minimal increase in propagation time with DBPF highlights its efficiency and scalability. Additionally, DBPF maintained the lowest number of exchanged blocks as there is no any redundant block sent within the propagation process, indicating superior bandwidth utilization and network performance.

Based on the results in Table I, the insights derived from the experimental results underscore the superior performance of the Dynamic Block Propagation Framework (DBPF) in terms of propagation time and bandwidth efficiency compared to DONS, RTT, and RNS. As the number of nodes increased from 50 to 450, the propagation time for DBPF showed a minimal increase, highlighting its scalability and efficiency. For instance, DBPF's propagation time increased by only 220.53 ms (from 105.89 ms to 326.42 ms) for networks with seven neighbors per node, whereas RTT's time surged by 32680.82 ms (from 2124.58 ms to 34805.4 ms), and RNS's time skyrocketed by 155497.16 ms (from 7489.9 ms to 162987.06 ms). This demonstrates DBPF's superior scalability and efficiency in larger networks.

Compared to DONS, DBPF also showed significant improvements. For networks with seven neighbors per node, DBPF reduced the propagation time by 60.61% for 50 nodes, 63.52% for 100 nodes, and 70.05% for 450 nodes. Similar trends were observed for networks with 15 neighbors per node. This significant reduction in propagation time showcases the effectiveness of DBPF in optimizing network performance and achieving fast propagation of new blocks to all nodes.

The performance of the proposed DBPF algorithm was compared with DONS, RTT, and RNS algorithms across two experimental setups. The total propagation time for each algorithm as the number of nodes increases is shown in Fig. 5.

In Experiment 1 (Fig. 5a), with an average of seven neighbors per node, DBPF significantly outperforms the other algorithms. For instance, with 450 nodes, DBPF achieves a propagation time of 326.42 ms, whereas DONS, RTT, and RNS require 1089.54 ms, 34805.4 ms, and 162987.06 ms, respectively.

In Experiment 2 (Fig. 5b), with an average of 15 neighbors per node, DBPF again demonstrates superior performance. With 450 nodes, DBPF has a propagation time of 156.95 ms, compared to DONS at 840.50 ms, RTT at 19918.79 ms, and RNS at 125417.10 ms.

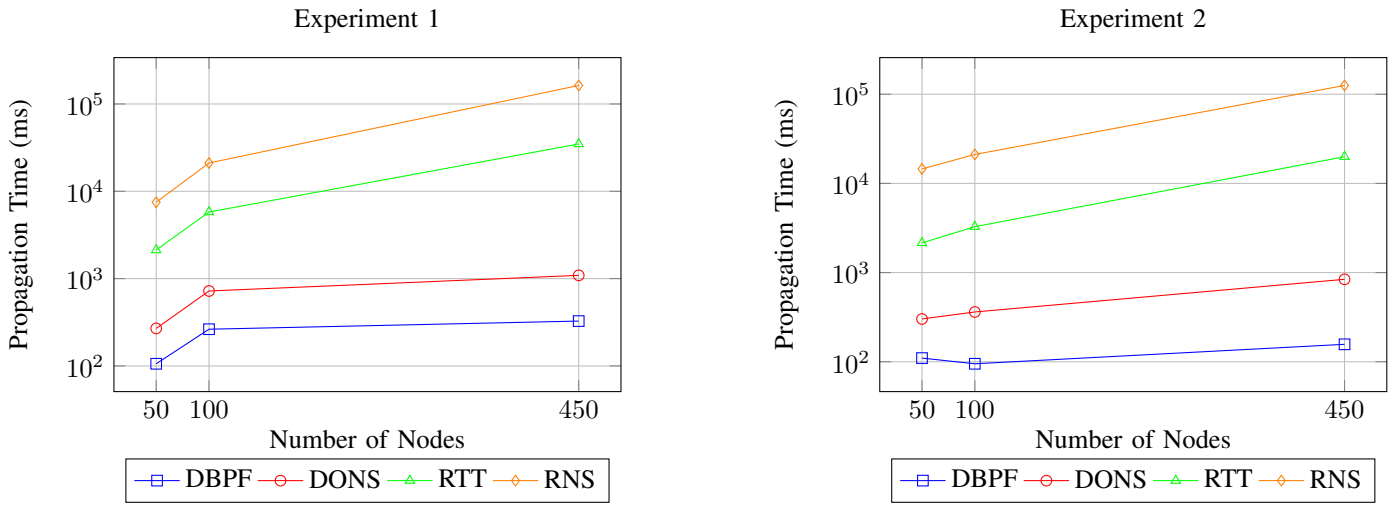
Experiment 2:

This experiment aims to demonstrate the robustness and efficiency of the Dynamic Block Propagation Framework (DBPF) under varying network bandwidth conditions, simulating realistic blockchain network scenarios under the Barabási-Albert (BA) model. Unlike previous experiments with constant bandwidth, this experiment introduces random bandwidth variations between nodes to assess DBPF's performance in more dynamic environments. The network configurations for this experiment include two cycles: Cycle 1 with an average number of neighbors per node set to 8 and Cycle 2 with an average number of neighbors per node set to 15. For each cycle, three different network sizes are tested: 30, 70, and 180 nodes. For each configuration, a random network topology is generated, and random bandwidth values are assigned to each pair of nodes to mimic real-world conditions where network bandwidth can vary. The total propagation time (PT) and the number of exchanged blocks (NB) are measured for DBPF, DONS, RTT, and RNS algorithms. The primary objective is to verify DBPF's efficiency in environments with fluctuating bandwidth and to compare its performance against other well-known algorithms (DONS, RTT, RNS). By simulating realistic network conditions, we aim to highlight DBPF's adaptability and robustness in maintaining low propagation times and efficient block propagation, even when network bandwidth varies significantly. The results from these experiments are presented in Table II.

The results of the second experiment, conducted under varied bandwidth conditions, provide a comprehensive evaluation of the performance of the Dynamic Block Propagation Framework (DBPF) compared to DONS, RTT, and RNS. The experiments were conducted on two different average neighbor settings (8 and 15) with randomly generated bandwidth values to simulate realistic network conditions. The bandwidth values for the first set of experiments ranged from 1000 to 10000 Bps, while the second set ranged from 5000 to 100000 Bps.

The experiments clearly demonstrate the impact of bandwidth variability on propagation time across different algorithms. In scenarios with lower average neighbors (8), DBPF consistently outperformed other methods. For instance, with 70 nodes, DBPF achieved a propagation time of 140.46 ms compared to DONS (647.99 ms), RTT (5406.97 ms), and RNS (16553.99 ms). This indicates that DBPF is highly efficient in utilizing available bandwidth, leading to reduced propagation times.

In scenarios with higher average neighbors (15), DBPF continued to show superior performance. With 30 nodes, DBPF recorded a propagation time of 93.91 ms, significantly lower than DONS (154.92 ms), RTT (1121.16 ms), and RNS (4998.62 ms). This trend persisted with 70 nodes and 180 nodes, where DBPF consistently demonstrated lower propagation times, highlighting its robustness in handling varying network conditions and higher node densities.



(a) Avg. Neighbors = 7

(b) Avg. Neighbors = 15

Fig. 5. Comparison of total propagation time for different algorithms in two experimental setups

TABLE II. PERFORMANCE OF DBPF, DONS, RTT, AND RNS UNDER VARIED BANDWIDTH CONDITIONS (BA MODEL)

Model	Nodes	Avg. Neighbors	Total Propagation Time (ms)			
			DBPF	DONS	RTT	RNS
BA	30	8	160.26	630.43	1609.75	10345.17
BA	70	8	140.46	647.99	5406.97	16553.99
BA	180	8	219.80	865.98	18618.33	63783.90
BA	30	15	93.91	154.92	1121.16	4998.62
BA	70	15	111.12	167.33	2084.82	10820.95
BA	180	15	130.91	217.12	5694.26	44358.25

To quantitatively assess the performance improvement of DBPF, we calculate the percentage reduction in propagation time compared to DONS. For instance, with 180 nodes and 8 average neighbors, DBPF achieved a propagation time of 219.80 ms, while DONS recorded 865.98 ms. The percentage improvement is calculated as follows:

$$\text{Percentage Improvement} = \left(\frac{\text{DONS Time} - \text{DBPF Time}}{\text{DONS Time}}\right) \times 100$$

$$\text{Percentage Improvement} = \left(\frac{865.98 - 219.80}{865.98}\right) \times 100 \approx 74.61\%$$

Similarly, with 30 nodes and 15 average neighbors, DBPF showed a 39.38% improvement over DONS. These substantial improvements demonstrate the efficiency of DBPF in minimizing propagation time, and enhancing overall network performance.

Let T_{DBPF} and T_{DONS} represent the propagation times of DBPF and DONS, respectively. The percentage improvement $P_{\text{improvement}}$ can be expressed as:

$$P_{\text{improvement}} = \left(\frac{T_{\text{DONS}} - T_{\text{DBPF}}}{T_{\text{DONS}}}\right) \times 100$$

The experiments also showed significant differences in the number of exchanged blocks (NB) across different methods. For instance, DBPF required no redundant block exchanges compared to RTT and RNS, reflecting its efficiency in reducing network load and overhead

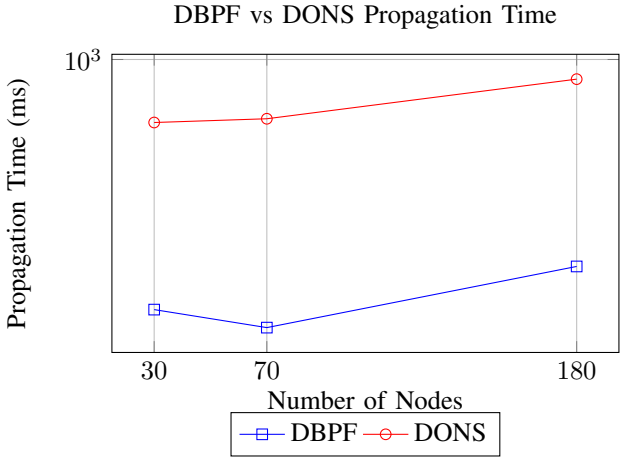


Fig. 6. Comparison of propagation time between DBPF and DONS.

These results underscore the effectiveness of DBPF in optimizing network performance, even under varying and challenging bandwidth conditions. The framework's ability to reduce propagation time significantly while maintaining low overhead highlights its potential for enhancing the scalability and efficiency of blockchain networks. As shown in Fig. 6, the DBPF algorithm consistently outperforms DONS across various network sizes and bandwidth conditions.

Experiment 3:

In this experiment, we investigate the efficiency of the Dynamic Block Propagation Framework (DBPF) implemented with different compression algorithms, specifically focusing on Brotli compression. The goal is to evaluate the performance of DBPF using Brotli in comparison with other widely-used compression techniques, zlib [29] and zstandard [30], under identical network conditions and the same block size. The Barabási-Albert (BA) model is employed to simulate realistic network environments.

The experiment consists of two cycles. In the first cycle, the number of nodes is set to 100 and 280, with an average number of neighbors per node fixed at 18. The bandwidth is kept constant at 10,000 bps to maintain a controlled environment. In the second cycle, the number of nodes is increased to 280 and 450, with the same average number of neighbors per node, but the bandwidth is randomly generated within the range of 1,000 to 10,000 bps to simulate dynamic network conditions.

The primary objective is to compare the performance of DBPF using Brotli, zlib, and zstandard compression algorithms. Zlib, a widely used compression library, provides a good balance between speed and compression ratio and is often utilized for data storage and transmission. Zstandard, developed by Facebook, offers high compression efficiency and fast decompression speeds, making it ideal for large-scale data compression tasks.

Metrics used to evaluate performance include total propagation time (ms). The hypothesis is that Brotli compression will outperform zlib and zstandard in reducing propagation time and optimizing network efficiency, demonstrating the robustness and scalability of the DBPF framework. The results from this experiment aim to provide a comprehensive analysis of the most efficient compression technique for enhancing blockchain network performance.

As it is shown in Table III, the results of this experiment highlight the effectiveness of the DBPF framework, particularly when using the Brotli compression algorithm. The propagation times for DBPF-Brotli were consistently lower than those for DONS and the other compression algorithms across all tested configurations. For instance, with 100 nodes and a fixed bandwidth of 10,000 bps, DBPF-Brotli achieved a propagation time of 105.65 ms, significantly outperforming DONS, which recorded 496.73 ms. Even when compared with other compression algorithms, DBPF-Brotli maintained its superiority, with DBPF-Zlib at 107.32 ms and DBPF-Zstandard at 216.80 ms. As the number of nodes increased to 280, DBPF-Brotli continued to show excellent performance, recording a propagation time of 91.45 ms, while DONS and DBPF-Zstandard recorded 627.63 ms and 297.06 ms, respectively. This trend was also observed when the bandwidth was randomly generated within the range of 1,000 to 10,000 bps. With 280 nodes, DBPF-Brotli achieved a propagation time of 163.32 ms, compared to DONS at 1496.8 ms, DBPF-Zlib at 171.27 ms, and DBPF-Zstandard at 189.71 ms. When examining the performance with 450 nodes under random bandwidth conditions, DBPF-Brotli again demonstrated superior performance with a propagation time of 99.65 ms. In contrast, DONS recorded 842.38 ms, DBPF-Zlib recorded 101.23 ms, and DBPF-Zstandard recorded 105.75 ms.

VI. Conclusion

The Dynamic Block Propagation Framework (DBPF) introduced in this study addresses key challenges in blockchain networks, such as limited transaction throughput, large blockchain sizes, scalability issues, and consensus protocol limitations. By integrating optimal neighbor selection (ONS) and advanced data compression techniques, specifically utilizing the Minimum Spanning Tree (MST) computation for efficient communication paths and the Brotli compression algorithm for data reduction, DBPF significantly enhances the performance of blockchain networks. Extensive experiments conducted using the Barabási-Albert (BA) model under various network conditions demonstrated the superiority of DBPF over existing methods. The outcomes of the presented research demonstrated a notable enhancement in block propagation across networks of diverse sizes, outperforming current state-of-the-art approaches. In networks with 180 nodes and 8 average neighbors, DBPF achieved up to a 74.61% improvement in propagation time compared to DONS. Further, DBPF outperformed RTT and RNS by 88.20% and 99.65%, respectively, under the same conditions. These results highlight the effectiveness of DBPF in reducing propagation time and enhancing overall network throughput. Additional experiments under varied bandwidth conditions confirmed the robustness and adaptability of DBPF. The framework maintained superior performance with both fixed and randomly generated bandwidths, demonstrating its flexibility in real-world scenarios. Evaluations with different compression algorithms, including Brotli, zlib, and zstandard, further underscored the efficiency of DBPF. Notably, DBPF using Brotli consistently outperformed other methods, validating the benefits of combining optimal neighbor selection with advanced compression techniques.

In summary, the DBPF framework offers a scalable and efficient solution for blockchain networks, significantly reducing block propagation time and increasing network throughput. These findings underscore the potential of DBPF to address inherent scalability challenges in blockchain technology, paving the way for more robust and efficient blockchain operations.

While the current study demonstrates the effectiveness of DBPF, several areas for future research can further enhance the framework's capabilities. Firstly, exploring the integration of DBPF with emerging consensus algorithms such as Proof of Stake (PoS) or Delegated Proof of Stake (DPoS) could provide insights into optimizing blockchain performance under different consensus mechanisms. future work could focus on implementing adaptive compression techniques that dynamically select the most suitable algorithm based on real-time network conditions and data characteristics.

References

[1] S. Nakamoto and A. Bitcoin, "A peer-to-peer electronic cash system," *Bitcoin.–URL: https://bitcoin. org/bitcoin. pdf*, vol. 4, no. 2, p. 15, 2008.

[2] O. Akanfe, D. Lawong, and H. R. Rao, "Blockchain technology and privacy regulation: Reviewing frictions and synthesizing opportunities," *International Journal of Information Management*, vol. 76, p. 102753, 2024.

[3] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum project yellow paper*, vol. 151, 2014.

[4] J. Liu and J. Wu, "A comprehensive survey on blockchain technology and its applications," *Highlights in Science, Engineering and Technology*, vol. 85, pp. 128–138, 2024.

TABLE III. PERFORMANCE OF DBPF WITH DIFFERENT COMPRESSION ALGORITHMS

Nodes	Avg. Neighbors	Total Propagation Time (ms)			
		DBPF (Brotli)	**DONS**	**DBPF (Zlib)**	**DBPF (Zstandard)**
100	18	105.65	496.73	107.32	216.80
280	18	91.45	627.63	93.75	297.06
280	18 (Random BW)	163.32	1496.8	171.27	189.71
450	18 (Random BW)	99.65	842.38	101.23	105.75

[5] P. Danzi, A. E. Kalør, Č. Stefanović, and P. Popovski, "Delay and communication tradeoffs for blockchain systems with lightweight iot clients," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2354–2365, 2019.

[6] W. A. Al-Nbhany, A. T. Zahary, and A. A. Al-Shargabi, "Blockchain-iot healthcare applications and trends: a review," *IEEE Access*, 2024.

[7] I. Mistry, S. Tanwar, S. Tyagi, and N. Kumar, "Blockchain for 5g-enabled iot for industrial automation: A systematic review, solutions, and challenges," *Mechanical systems and signal processing*, vol. 135, p. 106382, 2020.

[8] D. Das, S. Banerjee, K. Dasgupta, P. Chatterjee, U. Ghosh, and U. Biswas, "Blockchain enabled sdn framework for security management in 5g applications," in *Proceedings of the 24th International Conference on Distributed Computing and Networking*, pp. 414–419, 2023.

[9] S. Onopa and Z. Kotulski, "State-of-the-art and new challenges in 5g networks with blockchain technology," *Electronics*, vol. 13, no. 5, p. 974, 2024.

[10] A. K. Tyagi and S. Tiwari, "The future of artificial intelligence in blockchain applications," in *Machine learning algorithms using scikit and tensorflow environments*, pp. 346–373, IGI Global, 2024.

[11] A. M. S. Saleh, "Blockchain for secure and decentralized artificial intelligence in cybersecurity: A comprehensive review," *Blockchain: Research and Applications*, p. 100193, 2024.

[12] A. I. Sanka and R. C. Cheung, "A systematic review of blockchain scalability: Issues, solutions, analysis and future research," *Journal of Network and Computer Applications*, vol. 195, p. 103232, 2021.

[13] K. Croman, C. Decker, I. Eyal, A. E. Gencer, A. Juels, A. Kosba, A. Miller, P. Saxena, E. Shi, E. G. Sirer, *et al.*, "On scaling decentralized blockchains," in *International Conference on Financial Cryptography and Data Security*, pp. 106–125, Springer, 2016.

[14] G. Danner, I. Hegedűs, and M. Jelasity, "Improving gossip learning via limited model merging," in *International Conference on Computational Collective Intelligence*, pp. 351–363, Springer, 2023.

[15] W. Bi, H. Yang, and M. Zheng, "An accelerated method for message propagation in blockchain networks," *arXiv preprint arXiv:1809.00455*, 2018.

[16] M. A. Imtiaz, D. Starobinski, and A. Trachtenberg, "Empirical comparison of block relay protocols," *IEEE Transactions on Network and Service Management*, vol. 19, no. 4, pp. 3960–3974, 2022.

[17] Q. Zhou, H. Huang, Z. Zheng, and J. Bian, "Solutions to scalability of blockchain: A survey," *IEEE Access*, vol. 8, pp. 16440–16455, 2020.

[18] K. Ayinala, B.-Y. Choi, and S. Song, "Pichu: Accelerating block broadcasting in blockchain networks with pipelining and chunking," in *2020 IEEE International Conference on Blockchain (Blockchain)*, pp. 221–228, 2020.

[19] H. Baniata, A. Anaqreh, and A. Kertesz, "Dons: Dynamic optimized neighbor selection for smart blockchain networks," *Future Generation Computer Systems*, vol. 130, pp. 75–90, 2022.

[20] X. Zhang, W. Xia, X. Wang, J. Liu, Q. Cui, X. Tao, and R. P. Liu, "The block propagation in blockchain-based vehicular networks," *IEEE Internet of Things Journal*, vol. 9, no. 11, pp. 8001–8011, 2022.

[21] K. Wang and H. S. Kim, "Fastchain: Scaling blockchain system with informed neighbor selection," in *2019 IEEE International Conference on Blockchain (Blockchain)*, pp. 376–383, 2019.

[22] Y. Aoki and K. Shudo, "Proximity neighbor selection in blockchain networks," in *2019 IEEE International Conference on Blockchain (Blockchain)*, pp. 52–58, IEEE, 2019.

[23] C. Santiago and C. Lee, "Accelerating message propagation in blockchain networks," in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 157–160, IEEE, 2020.

[24] J. Wang, "Txilm: Lossy block compression with salted short hashing," June 21 2022. US Patent 11,368,286.

[25] R. Lewis, *Guide to graph colouring*. Springer, 2021.

[26] P. Ayegba, J. Ayoola, E. Asani, and A. Okeyinka, "A comparative study of minimal spanning tree algorithms," in *2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS)*, pp. 1–4, IEEE, 2020.

[27] M. Pasetti, E. Sisinni, P. Ferrari, P. Bellagente, and D. Zaninelli, "Comprehensive evaluation of lossless compression algorithms in a real use case for smart grid applications," *Sustainable Energy, Grids and Networks*, vol. 36, p. 101238, 2023.

[28] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.

[29] H. Yang, G. Qin, and Y. Hu, "Compression performance analysis of different file formats," *arXiv preprint arXiv:2308.12275*, 2023.

[30] Y. Collet, "Rfc 8878: Zstandard compression and the'application/zstd'media type," 2021.

Skin Diseases Classification with Machine Learning and Deep Learning Techniques: A Systematic Review

Amina Aboulmira[1], Hamid Hrimech[2], Mohamed Lachgar[3]
LAMSAD Laboratory, ENSA, Hassan First University, Berrechid, Morocco[1,2]
LTI Laboratory, ENSA, Chouaib Doukkali University, El Jadida, Morocco[3]
L2IS Laboratory, Faculty of Science and Technology, University Cadi Ayyad, Marrakech, Morocco[3]
Higher Normal School, Department of Computer Science, University Cadi Ayyad, Marrakech, Morocco[3]

***Abstract*—Skin cancer is one of the most prevalent types of cancer worldwide, and its early detection is crucial for improving patient outcomes. Artificial Intelligence (AI) has shown significant promise in assisting dermatologists with accurate and efficient diagnosis through automated skin disease classification. This systematic review aims to provide a comprehensive overview of the various AI techniques employed for skin disease classification, focusing on their effectiveness across different datasets and methodologies. A total of 220 articles were initially identified from databases such as Scopus and IEEE Xplore. After removing duplicates and conducting a title and abstract screening, 213 studies were assessed for eligibility based on predefined criteria such as study relevance, clarity of results, and innovative AI approaches. Following full-text review, 56 studies were included in the final analysis. These studies were categorized based on the AI techniques used, including Convolutional Neural Networks (CNNs), Transformer-based models, hybrid models combining CNNs with other techniques, Generative Adversarial Networks (GANs), and ensemble learning approaches. The review highlights that the ISIC dataset and its variations are the most commonly used data sources, owing to their extensive and diverse collection of dermoscopic images. The results indicate that CNN-based models remain the most widely adopted and effective approach for skin disease classification, with several hybrid and Transformer-based models also demonstrating high accuracy and specificity. Despite the advancements, challenges such as dataset variability, the need for more diverse training data, and the lack of interpretability in AI models persist. This review provides insights into current trends and identifies future directions for research, emphasizing the importance of integrating AI into clinical practice for improved skin disease management.**

Keywords*—Skin Disease Classification; Artificial Intelligence (AI); Convolutional Neural Networks (CNNs); Transformer-based Models; Generative Adversarial Networks (GANs); ensemble learning; hybrid models; ISIC dataset; dermatology; machine learning; deep learning; skin cancer detection; dermoscopic images; medical imaging; systematic review

I. Introduction

Skin diseases encompass a broad spectrum of conditions that affect the skin, which is the largest organ of the human body. These conditions can vary greatly in severity and presentation, ranging from benign issues like acne to life-threatening diseases such as melanoma. Accurate diagnosis, classification, and segmentation of skin diseases are critical as they directly influence treatment plans, patient outcomes, and overall healthcare efficiency. Misdiagnosis or delayed diagnosis can lead to severe consequences, including unnecessary treatments or the progression of the disease to more advanced stages [1], [2].

Common skin diseases that are frequently the focus of classification and segmentation efforts in medical research include melanoma, psoriasis, and eczema. Melanoma, in particular, is a malignant tumor of melanocytes and is one of the most serious forms of skin cancer due to its high potential for metastasis [3]. Psoriasis is a chronic inflammatory skin condition characterized by the rapid growth of skin cells, leading to thick, red, scaly patches [4]. Eczema, also known as atopic dermatitis, is a chronic condition that causes inflamed, itchy, cracked, and rough skin [5].

Traditionally, the diagnosis of skin diseases has relied heavily on clinical examinations performed by dermatologists. This process typically involves visual inspection, often aided by tools like dermoscopy, which provides magnified images of the skin, allowing for better visualization of structures beneath the skin surface [6]. In cases where the visual inspection is inconclusive, a biopsy may be performed, wherein a sample of the skin is taken for histopathological examination under a microscope [7].

While these traditional methods are well-established and widely used, they are not without limitations. Human error is a significant concern, as the accuracy of diagnosis can vary depending on the dermatologist's experience and expertise. Studies have shown variability in diagnostic accuracy, even among experienced dermatologists [8]. Additionally, the manual segmentation of skin lesions, which is crucial for treatment planning, is time-consuming and labor-intensive. This process often involves delineating the borders of the lesion manually, which is not only subjective but also prone to variability [9].

Given these challenges, there is a growing interest in the application of artificial intelligence (AI) to improve the accuracy, efficiency, and consistency of skin disease diagnosis and segmentation. A significant body of research has emerged that integrates AI techniques into the detection and segmentation of skin diseases, demonstrating promising results in enhancing diagnostic processes. This paper proposes a systematic review focusing on the use of AI for the classification and segmentation of skin diseases, aiming to consolidate current findings, identify prevailing methodologies, and highlight existing knowledge gaps within this rapidly evolving field. By systematically analyzing the literature, this review will provide

valuable insights into the effectiveness of AI in dermatology, outline the strengths and limitations of current approaches, and suggest potential avenues for future research.

The remainder of this paper is organized as follows: The Theoretical Background section provides a comprehensive overview of the foundational concepts and existing literature related to AI-driven skin disease classification. This is followed by the Methodology section, where the criteria and procedures used to select and analyze the studies are detailed. The Results section presents the findings from the systematic review, including bibliometric analyses and the evaluation of methodologies. In the Discussion section, the implications of the findings are explored, and the strengths and limitations of current approaches are assessed. Finally, the paper concludes with Final Considerations, summarizing the key takeaways and suggesting directions for future research in this rapidly evolving field.

II. Theoretical Background

A. From Classic to Modern Dermatology

Dermatology has undergone a profound transformation from its early days of visual inspection and basic histopathological analysis to the incorporation of advanced imaging technologies and digital tools. Traditionally, dermatologists relied heavily on their clinical expertise, using simple tools like magnifying glasses to examine skin lesions, and performing biopsies followed by histopathological analysis to diagnose complex conditions. Histopathological images, derived from these biopsies, provided detailed views of skin tissue at the cellular level, becoming a cornerstone of accurate diagnosis, particularly for skin cancers [10].

Over time, the limitations of these conventional methods, including their invasiveness and the potential for diagnostic variability, drove the development of more sophisticated tools. The advent of dermoscopy revolutionized non-invasive examination by enabling magnified visualization of subsurface skin structures, greatly enhancing the accuracy of initial assessments [11]. This was followed by the introduction of digital dermoscopy, confocal microscopy, optical coherence tomography (OCT), and multispectral imaging, each offering unique insights into different aspects of skin anatomy and pathology [12], [13]. These technological advancements have not only improved traditional diagnostic practices but have also set the stage for the integration of artificial intelligence, which leverages these diverse imaging modalities to further revolutionize dermatological care.

These diverse imaging modalities—ranging from high-resolution dermoscopic images to detailed histopathological and confocal microscopy images—have provided a wealth of data that is now being harnessed by artificial intelligence (AI) to further advance the field. AI systems, trained on these extensive image datasets, are capable of analyzing and interpreting complex patterns within the skin, leading to more precise and efficient diagnoses [14], [15]. This integration of AI with a wide variety of imaging techniques is not only enhancing diagnostic performance but also paving the way for more personalized and effective treatment strategies in dermatology.

B. Artificial Intelligence and Skin Disease Classification

Building on the advancements in imaging technologies, the integration of artificial intelligence into dermatology marks a significant leap forward in skin disease classification. AI, through machine learning and deep learning algorithms, leverages the vast array of imaging data—from clinical and dermoscopic images to histopathological and confocal microscopy images—to identify subtle patterns and features that might escape human detection. This capability allows AI to offer unprecedented accuracy and efficiency in the diagnosis and classification of skin conditions, setting the stage for more personalized and precise dermatological care [16].

Artificial Intelligence (AI) is a broad field that includes machine learning (ML) and deep learning (DL), both of which have become crucial in healthcare, particularly in dermatology. ML involves algorithms that learn from data to make predictions, while DL, a more advanced subset, uses multi-layered neural networks to automatically extract features from complex datasets. These AI techniques are particularly well-suited for analyzing skin images, enabling more accurate and efficient diagnosis of dermatological conditions by identifying patterns that may be difficult for human clinicians to detect [17], [14], [16].

Machine learning (ML) initially made significant strides in dermatology by enabling the automated classification of skin lesions based on manually extracted features. Techniques such as support vector machines (SVMs), decision trees, and random forests were effectively used to analyze images and identify patterns that distinguish between benign and malignant conditions [14]. One notable success was the use of SVMs in melanoma detection, where these models achieved high accuracy by focusing on key features like color, texture, and shape [16]. Despite these advancements, ML models often required extensive feature engineering, relying on expert knowledge to select the most relevant attributes. This limitation, combined with the models' relatively shallow architecture, made it challenging to handle the complex and high-dimensional data typical in dermatology, which paved the way for the adoption of deep learning (DL).

Machine learning was the first wave of AI to make a substantial impact in dermatology. By training algorithms on datasets of labeled skin images, ML models have been able to assist in diagnosing various skin conditions. For example, support vector machines (SVMs), decision trees, and ensemble methods like random forests have been utilized to classify skin lesions based on features extracted manually or through basic automated processes [14]. These models improved diagnostic consistency and reduced human error, particularly in distinguishing between benign and malignant lesions. However, the effectiveness of ML in dermatology was often limited by the need for extensive feature engineering and the relatively shallow nature of these models, which struggled to capture the complexity of high-dimensional image data.

The introduction of deep learning marked a significant leap forward for AI in dermatology. Deep learning, particularly through convolutional neural networks (CNNs), overcame many of the limitations of traditional ML models by automatically learning hierarchical features directly from raw image data. CNNs, with their ability to process and analyze

large amounts of image data, have been particularly effective in dermatology, where they are used to classify skin diseases with unprecedented accuracy [17]. These models have been trained on vast datasets of clinical, dermoscopic, and histopathological images, enabling them to recognize subtle patterns and features that might be missed by human clinicians or simpler algorithms.

As AI continues to evolve, its applications in dermatology are expected to expand, addressing current challenges such as model interpretability and data bias, while opening new avenues for personalized and accessible skin care.

III. Related Work

The application of machine learning (ML) and deep learning (DL) techniques in the classification of skin diseases has garnered significant attention in recent years, leading to the development of various models aimed at improving diagnostic accuracy and efficiency. Traditional ML methods such as Support Vector Machines (SVM), Random Forests (RF), and k-Nearest Neighbors (k-NN) have been extensively utilized for skin disease classification tasks. For instance, [14] employed SVM in conjunction with handcrafted features for melanoma detection, achieving competitive performance against dermatologists. However, these approaches are often limited by their dependence on feature engineering, which can be both time-consuming and reliant on domain expertise.

With the advent of DL, particularly Convolutional Neural Networks (CNNs), there has been a paradigm shift in the approach to skin disease classification. CNNs have the inherent ability to automatically learn hierarchical feature representations from raw image data, eliminating the need for manual feature extraction.

Over the past decade, several systematic reviews have been conducted to assess the efficacy of machine learning (ML) and deep learning (DL) techniques in the classification of skin diseases. These reviews have provided valuable insights into the trends, challenges, and future directions in this rapidly evolving field. The author in [18] conducted one of the early comprehensive reviews, focusing on the application of CNNs in dermatology. This review highlighted the increasing adoption of deep learning models over traditional machine learning approaches due to their superior accuracy and ability to process raw image data without extensive preprocessing.

Further advancing the field, [19] provided a thorough review of dermatological image analysis using both machine learning and deep learning techniques. The review emphasized the significant advancements in CNN architectures, such as ResNet, AlexNet, VGG.., and their impact on improving diagnostic accuracy for various skin conditions, including common cases and rare ones. However, the authors also noted the limitations related to the interpretability of these models and the challenges posed by imbalanced datasets.

[20] conducted a meta-analysis focused specifically on the performance comparison between human dermatologists and DL models. Their review concluded that DL models, particularly those based on CNNs, have reached a level of performance comparable to that of expert dermatologists, especially in the detection of malignant melanoma. This finding was corroborated by the review conducted by [21], which compared multiple DL models and found that ensemble methods often outperform individual models in terms of accuracy and robustness.

A more recent review by [22] explored the integration of advanced techniques such as transfer learning and generative adversarial networks (GANs) into dermatological applications. The authors highlighted that while these techniques offer promising avenues to overcome the challenges of limited labeled data and improve model generalizability, there is still a need for more standardized evaluation protocols and larger, more diverse datasets to fully realize their potential.

In addition to these, [23] reviewed the ethical and regulatory considerations associated with the deployment of ML and DL models in clinical settings. Their work underscores the importance of ensuring model transparency, patient data privacy, and the need for rigorous clinical validation before these models can be widely adopted in practice.

These reviews collectively illustrate the rapid advancements and the ongoing challenges in applying ML and DL techniques to skin disease classification. They provide a foundation for future research, particularly in addressing issues related to model interpretability, dataset bias, and the ethical implications of AI in dermatology.

In contrast, this systematic review aims to fill a specific gap in the literature by focusing on the progression from traditional ML techniques to advanced DL models. Process was to systematically search, extract, and analyze studies that detail the exact methodologies and techniques used in both ML and DL for skin disease classification. This approach allows us to map the evolution of these techniques, highlighting how deep learning, particularly CNNs, Transformers, and hybrid models, has been increasingly adopted and refined over time.

A notable particularity of this review is the attention given to hybrid models that combine both machine learning and deep learning techniques. This focus is crucial, as hybrid models represent a significant trend in the literature, often outperforming their pure ML or DL counterparts by leveraging the strengths of both approaches. These models' architectures and performances were meticulously documented, thereby offering insights into their potential for future research and clinical applications, by tracing the trajectory from early adoption phases to the more recent innovations, such as Vision Transformers (ViTs) and Generative Adversarial Networks (GANs), providing a clear picture of the technological advancements and their impact on classification performance.

This systematic review stands out in its detailed exploration of the transition from ML to DL in skin disease classification. It not only contributes to the existing body of knowledge but also serves as a valuable resource for researchers aiming to further advance the field.

IV. Methodology

A. Aim and Scope Definition

The primary objective of this study is to systematically review and synthesize the existing literature on the application of artificial intelligence (AI), including machine learning (ML)

and deep learning (DL), in the classification of skin diseases. The study aims to explore how these AI techniques have been utilized in analyzing various types of dermatological images, including clinical, dermoscopic, and histopathological images, to enhance diagnostic accuracy and efficiency. The scope of the review includes original research articles, and review papers published in peer-reviewed journals. The study focuses on skin diseases such as melanoma, psoriasis, and acne, with no specific geographical or linguistic restrictions, although only English-language publications are included.

B. Data Collection

1) Retrieval database: The literature for this systematic review was retrieved from several databases known for their comprehensive coverage of medical and technological research. The primary databases used include IEEE Xplore and Scopus. These databases were selected due to their relevance to the fields of dermatology, artificial intelligence, and medical imaging, ensuring that a broad range of studies could be captured.

2) Keyword selection strategy: A strategic keyword selection process was employed to identify relevant studies. Keywords were chosen to reflect the core concepts of the study: artificial intelligence, machine learning, deep learning, and skin disease classification. Specific search terms included combinations of the following: "AI in dermatology", "deep learning", "machine learning", "classification". Boolean operators (AND, OR, NOT) were used to refine the search and ensure comprehensive coverage of the literature.

Research query for scopus = (TITLE-ABS-KEY ("skin diseases" OR "dermatological disorders" OR "skin conditions") AND ("machine learning" OR "deep learning") AND ("classification") AND TITLE-ABS-KEY ("segmentation") AND PUBYEAR ¿ 2020 AND PUBYEAR ¡ 2024 AND [LIMIT-TO (LANGUAGE, "English")

Research query for IEEE Xplore = ("All Metadata": "skin diseases" OR "All Metadata": "dermatological disorders" OR "All Metadata": "skin conditions" OR "All Metadata": "skin lesions") AND ("All Metadata": "machine learning" OR "All Metadata": "deep learning") AND ("All Metadata": "classification") NOT ("All Metadata": "segmentation")] (Table I).

C. Inclusion and Exclusion Criteria

1) Inclusion criteria: To ensure the relevance and quality of the included studies, the following inclusion criteria were applied:

- The study focuses on the application of AI (ML or DL) in the classification of skin diseases.
- The study is published in a peer-reviewed journal.
- The study provides sufficient data for analysis, including details of the AI methods used and the types of images analyzed.
- The study is written in English.

TABLE I. RESEARCH QUESTIONS OF THE STUDY

No.	Research Question (RQ)
RQ1	What are the most commonly used artificial intelligence models for the classification of skin diseases?
RQ2	How effective are these AI models in accurately classifying various skin diseases?
RQ3	How do AI-based methods for skin disease classification compare to traditional diagnostic methods ?
RQ4	What datasets are commonly used to train and validate AI models for skin disease classification, and what are their characteristics?
RQ5	What are the potential clinical implications of integrating AI models into the diagnosis and treatment planning of skin diseases?
RQ6	What are the main challenges and limitations associated with the use of AI models in the classification of skin diseases?
RQ7	What future research directions are needed to improve the performance and clinical applicability of AI models in skin disease classification?

2) Exclusion criteria: Studies were excluded based on the following criteria:

- The study does not focus on dermatology or AI applications in skin disease classification.
- The study is a conference abstract, editorial, letter, or opinion piece with no original research data.
- The study lacks methodological rigor, as determined by the quality assessment process.
- The study does not utilize one of the following datasets: "ISIC", "HAM10000", "PH2", "Dermnet" or "Derm7pt".
- The study is about the segmentation of skin diseases.
- The study is not available in English.

In this systematic review, a total of 220 articles were initially identified through comprehensive database searches in Scopus and IEEE Xplore. After the removal of 7 duplicate articles, 213 unique records remained for screening.

The initial screening process led to no exclusions at this stage. All 213 records were then assessed for eligibility. Subsequently, 109 articles were excluded due to various reasons, such as lack of clear results reporting (11 articles), non-novel approaches (11 articles), irrelevant content (15 articles), or falling out of the scope of the review (14 articles). This left 104 articles for further retrieval attempts, out of which 48 were not retrieved. Ultimately, 56 articles were included in the final review and deemed suitable for qualitative synthesis (Fig. 1).

3) Software tools: The data analysis was performed using a combination of software tools, including Microsoft Excel for data management and descriptive analysis, R for statistical analysis, and Bibliometrix with Biblioshiny package for qualitative analysis. These tools were chosen based on their functionality, ease of use, and ability to handle large datasets effectively.

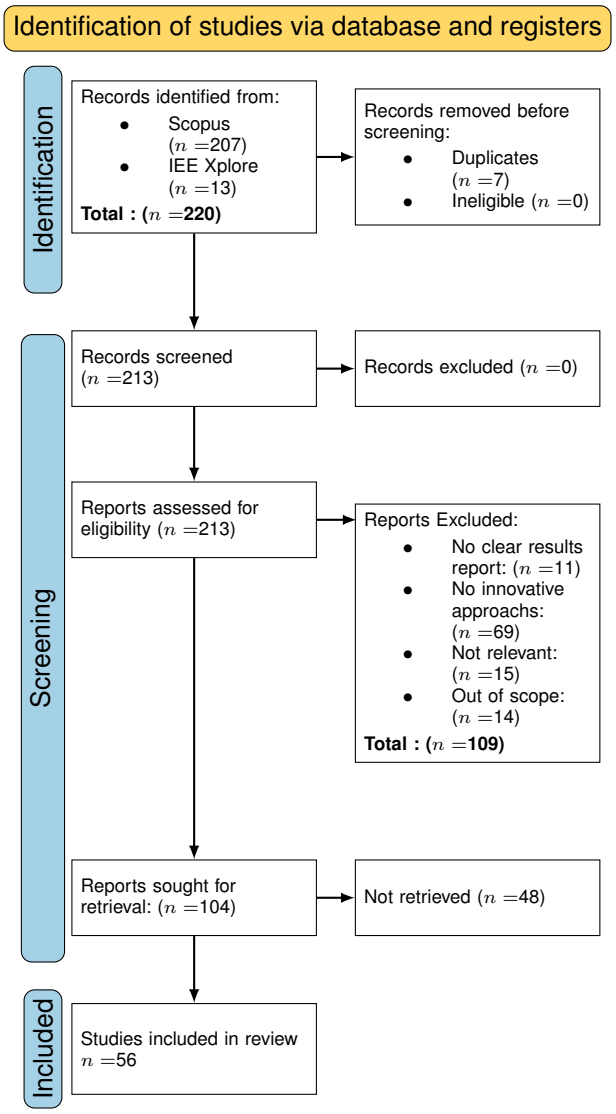


Fig. 1. Prisma diagram.

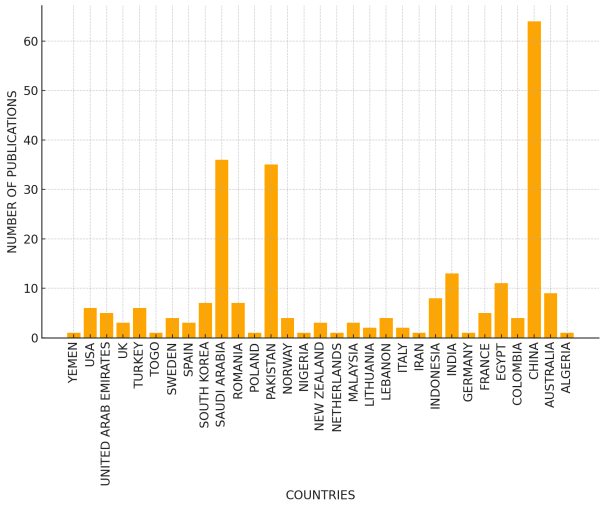


Fig. 2. Distribution of selected papers by countries.

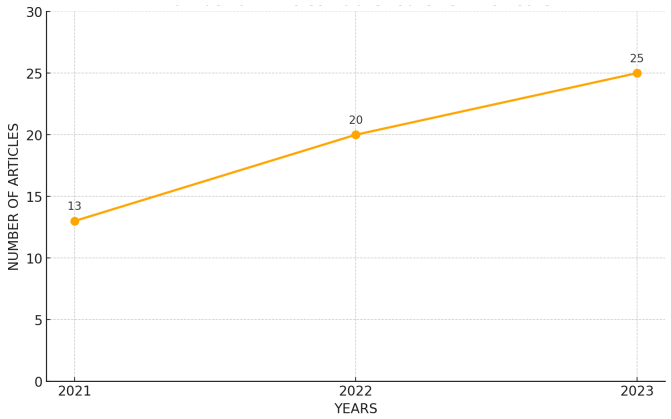


Fig. 3. Number of articles by years of publication.

V. Results

Retained articles related to the classification of skin diseases using machine learning and deep learning tools were published across 31 different countries.

Fig. 2 illustrates the distribution of included works by countries of publication. The countries representing the origin of the greatest number of publications are China with 64 articles, followed by Pakistan with 35 articles, and Saudi Arabia with 36 articles. Additionally, India and Indonesia contributed significantly with 13 and 8 articles, respectively. Together, China, Pakistan, Saudi Arabia, India, and Indonesia account for over 80% of the papers included in this study, with a total of 56 papers.

Considering publication dates of selected articles, a large part of the retained documents was published during the period 2020 to 2023, with a total of 45 papers. Fig. 3 below represents the curve of the publication's evolution per years between 2021 and 2023.

The bibliometric analysis also includes a thematic map in Fig. 4, which offers a visual representation of the key research themes within the field of AI-driven skin disease classification. The map categorizes themes based on their relevance (centrality) and development degree (density), providing a clear overview of the research landscape. This analysis highlights "deep learning" and "dermatology" as highly relevant but still developing themes, suggesting ongoing growth in these areas. Conversely, "human melanoma" appears as a well-established motor theme, indicating its foundational role in the field. Emerging or declining themes such as "optimization algorithms" are also identified, pointing to areas where future research may be necessary.

To further explore the thematic connections within the literature, Fig. 5 presents a co-occurrence network of keywords. This visualization highlights the central themes in AI-driven skin disease classification, with "deep learning" and "skin cancer" emerging as the most connected and frequent terms. The network also reveals the relationships between various methodologies, such as "convolutional neural networks" and "transfer learning", underscoring their significance in this research domain.

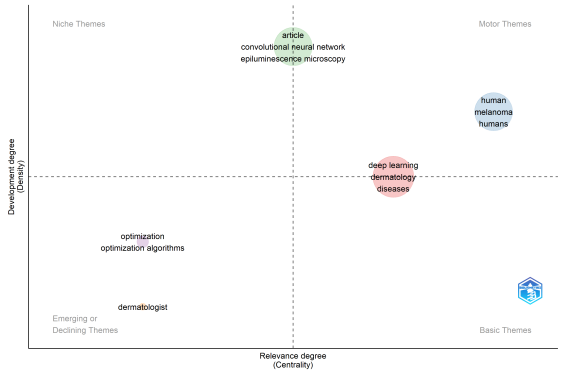


Fig. 4. Thematic map of research in AI-driven skin disease classification.

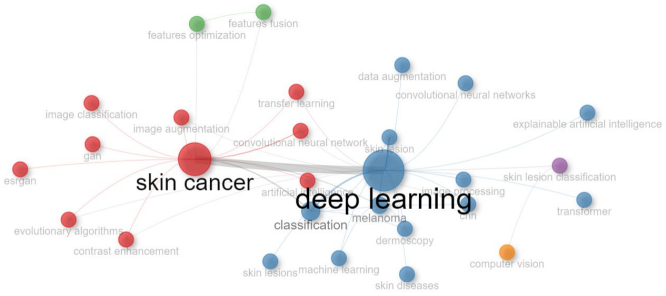


Fig. 5. Co-occurrence network of keywords in AI-driven skin disease classification literature. This network visualizes the relationships between key terms, with larger nodes indicating more frequently occurring keywords. The network highlights the centrality of terms like "deep learning" and "skin cancer", revealing their strong connections with related concepts such as "image augmentation", "transfer learning", and "convolutional neural networks".

RQ1: What are the most commonly used artificial intelligence models for the classification of skin diseases?

The qualitative analyse of the obtained results from the researches papers provided, indicates that Convolutional Neural Networks (CNNs) are the most frequently employed AI models for skin disease classification. CNNs appear in various forms, including standalone models, hybrid approaches (e.g. CNN combined with SVM, XGBoost, or Random Forest), and enhanced architectures (e.g. Inception-V3, EfficientNets). Additionally, Transformer-based models and Vision Transformers (ViTs) are increasingly utilized, reflecting a shift towards more complex, attention-based architectures. Hybrid models integrating multiple techniques, such as CNNs with transformers or Extreme Learning Machines (ELM), are also prevalent, suggesting an ongoing effort to enhance model performance through combining strengths of different methods.

RQ2: How effective are these AI models in accurately classifying various skin diseases?

The reviewed studies report high accuracy rates across different AI models, with CNN-based approaches often achieving accuracy levels above 90%. Specific examples include:

- A deep learning-based auto-encoder with an accuracy of 98.7%.
- A CNN combined with XGBoost and SVM showing 97.85% accuracy.
- Hybrid CNN-ELM models achieving up to 96.7% accuracy.

Sensitivity and specificity metrics are also robust, frequently exceeding 90%. For example, the sensitivity for CNN models ranges from 88.3% to 98.46%, while specificity ranges from 90% to 100%. These results indicate that AI models, particularly CNNs and hybrid approaches, are effective in accurately classifying skin diseases, often outperforming traditional diagnostic approaches.

RQ3: How do AI-based methods for skin disease classification compare to traditional diagnostic methods?

The analyse of the provided researches reveals that deep learning-based methods, particularly Convolutional Neural Networks (CNNs), consistently outperform traditional machine learning techniques in skin disease classification across various metrics. Deep learning models achieve higher accuracy, often exceeding 94%, with some models reaching up to 96.83% ([24], [25]). In contrast, traditional machine learning methods, including Support Vector Machines (SVM) and Random Forests, typically demonstrate accuracy within the 90% to 97% range.

Sensitivity and specificity are also higher in deep learning models. For instance, [26] reports a sensitivity of 98% and specificity of 98.1% using a Lightweight CNN with dynamic-sized kernels and ReLU/leaky ReLU activations, while machine learning approaches generally show slightly lower performance in these metrics.

Moreover, deep learning models benefit from data augmentation techniques, such as Generative Adversarial Networks (GANs), which enhance their generalization capabilities, especially in imbalanced datasets ([27]). Traditional machine learning methods, although effective, often require more extensive feature engineering and do not leverage data augmentation as effectively as deep learning approaches.

Hybrid models that combine deep learning with machine learning methods, such as CNNs with SVM or Random-Forest offer a balanced approach, leveraging the strengths of both techniques ([28], [29]). However, these hybrids still typically fall short of purely deep learning-based methods in terms of overall performance.

In terms of model interpretability, traditional machine learning models, particularly decision trees, provide more straightforward explanations. However, the integration of Explainable AI (XAI) techniques with CNNs has begun to address the "black box" nature of deep learning models, enhancing their transparency ([30]).

While machine learning models still hold value in scenarios requiring interpretability, deep learning approaches, particularly when augmented with hybrid techniques, represent the most effective tools for skin disease diagnostics.

RQ4: What datasets are commonly used to train and validate AI models for skin disease classification, and what are their characteristics?

Several datasets emerge as commonly utilized benchmarks for training and validation purposes for skin disease classification task. The following are the key datasets frequently employed in these studies, along with their characteristics:

- ISIC (International Skin Imaging Collaboration): This dataset is a comprehensive repository of dermoscopic images that has been widely adopted due to its diversity and scale [31]. The ISIC dataset includes several variations, such as ISIC2018, ISIC2019, ISIC2020, ISIC2017, and ISIC2008, each corresponding to different years of challenge submissions. These variations contain images that differ in terms of disease types, resolutions, and annotations, providing a robust foundation for model development (Fig. 6).
- HAM10000: The HAM10000 dataset [32] contains a large collection of dermatoscopic images with an emphasis on the most common pigmented skin lesions. It is particularly valued for its balanced representation of multiple classes of skin diseases, making it a reliable resource for training classifiers that need to generalize across various conditions.
- PH2: Although smaller, the PH2 dataset is a key resource that provides high-quality dermoscopic images specifically curated for the assessment of melanocytic and non-melanocytic skin lesions [33]. Its limited class diversity is counterbalanced by the precise annotations and image quality, making it ideal for specialized classification tasks.
- Derm7pt: This dataset focuses on a seven-point checklist system for melanoma detection, providing a structured approach to training models in clinical decision-making scenarios [34].
- PAD-UFES-20: This dataset includes images collected from a specific demographic, aiding in the development of models that are more inclusive and adaptable to different population groups [35].

Fig. 7 summarizes the usage distribution of these datasets across the reviewed studies, highlighting the dominance of the ISIC and its variants, followed by the widespread adoption of HAM10000 and other datasets.

RQ5: What are the potential clinical implications of integrating AI models into the diagnosis and treatment planning of skin diseases?

The high accuracy and specificity of AI models indicate significant potential for their integration into clinical practice. AI can assist dermatologists in diagnosing skin diseases more quickly and accurately, reducing the time required for analysis and potentially improving patient outcomes. Moreover, the precision of AI models could help in early detection of malignant lesions, leading to more timely interventions. However, the clinical integration of these models also requires careful consideration of ethical implications, including patient consent, data privacy, and the potential for algorithmic bias. The integration of AI into treatment planning could also extend to personalized medicine, where AI-driven insights help tailor treatment strategies to individual patient profiles.

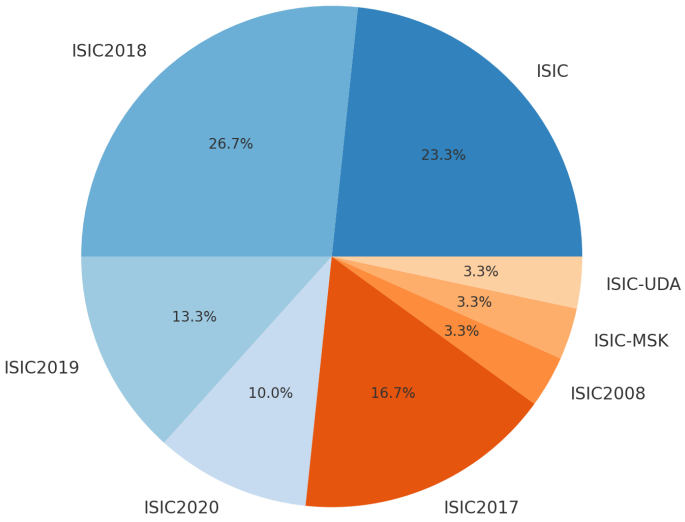


Fig. 6. Usage distribution of ISIC dataset variants across studies.

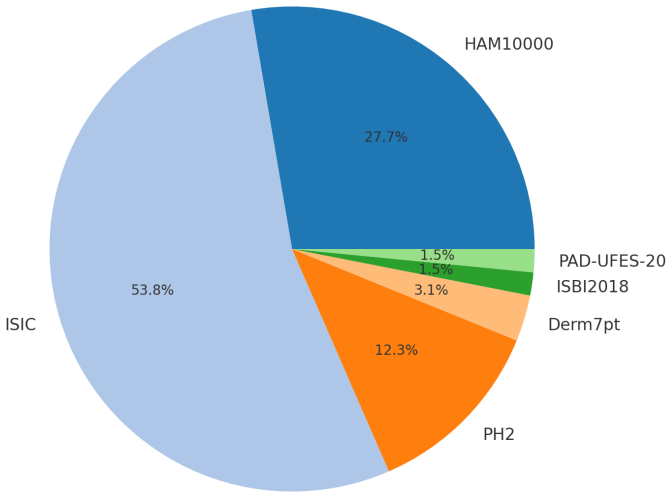


Fig. 7. Percentage of Dataset Usage in the Selected Studies.

RQ6: What are the main challenges and limitations associated with the use of AI models in the classification of skin diseases?

Despite the promising results, several challenges and limitations persist. One significant challenge is the generalizability of AI models across different populations and clinical settings. Most models are trained on specific datasets, which may not fully represent the diversity of skin types and conditions found in the broader population. Data quality and annotation is another concern; the accuracy of AI models heavily depends

on the quality of the input data, and misannotations or inconsistencies can lead to reduced model performance. Moreover, computational complexity and the need for high-end hardware for training and deploying complex models like transformers could limit their accessibility in resource-constrained environments. Additionally, there is a need for explainability in AI models to ensure that clinicians can trust and understand the decisions made by these systems. Finally, there is limited research on how these models can be effectively integrated into real-world clinical workflows. Understanding the challenges of AI adoption in everyday clinical practice, including clinician trust, usability, and workflow compatibility, is essential for the successful implementation of AI in dermatology [36].

RQ7: What future research directions are needed to improve the performance and clinical applicability of AI models in skin disease classification?

Future research in AI for skin disease classification should prioritize several key areas to enhance both performance and clinical applicability. First, increasing dataset diversity and size is critical for improving model generalization, particularly for underrepresented skin conditions and demographic groups, ensuring broader applicability across diverse populations. Second, cross-validation in varied clinical settings is necessary to assess the robustness and consistency of AI models, which is essential for their practical integration into healthcare environments. Third, improving the explainability of AI models is vital for building trust among healthcare providers, enabling transparent and verifiable diagnoses that can encourage clinical adoption.

Additionally, research should focus on reducing the computational complexity of AI models, making them more accessible and deployable, especially in resource-limited settings. The development of multi-modal diagnostic systems that integrate visual data with patient-specific information, such as demographics and medical history, is another crucial area. These systems could enhance diagnostic accuracy and lead to more personalized treatments.

Moreover, most current research focuses on static images and one-time predictions. There is a gap in studies that investigate the application of AI in longitudinal analyses, where the progression of skin diseases is tracked over time. Such studies are crucial for developing AI tools that can not only diagnose but also monitor disease progression and treatment response [37], [38].

VI. Discussion

The analysis of methodologies employed across the studies included in this systematic review reveals distinct patterns in the application of various AI techniques for skin disease classification. To provide a more nuanced understanding of the approaches utilized by different researchers, the studies have been categorized based on the general technique employed, including Machine Learning techniques (e.g. XGBoost, SVM), Generative Adversarial Networks (GANs), Ensemble Learning approaches, Multi-modal methods, Transformer-based models, and Hybrid models combining Convolutional Neural Networks (CNNs) with other techniques.

The following sections summarize the methodologies applied within each of these categories, highlighting the specific strategies used in feature extraction, model training, and validation. This organization elucidates the strengths and limitations inherent to each approach, offering a clearer perspective on the current state of AI-driven skin disease classification. By structuring the analysis in this manner, a direct comparison between different techniques is facilitated, providing insights into potential future directions for research in this rapidly evolving field.

A. Studies Using Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) have been explored for their ability to generate synthetic data and improve classification tasks in skin disease detection.

Two notable studies have employed Generative Adversarial Networks (GANs) to enhance skin lesion classification. Abdelhalim et al. [39] combined GANs with Convolutional Neural Networks (CNNs) in their study, utilizing the ISIC dataset. In this approach, GANs were employed to generate synthetic dermoscopic images, which were subsequently used to train the CNN. This method effectively improved the model's robustness by augmenting the training data.

Similarly, Su et al. [27] adopted a GAN-CNN hybrid model for skin lesion classification. In their study, the GAN was used to generate additional training samples, while the CNN handled the classification task. This methodology was validated using the HAM10000 dataset, and the results demonstrated improved accuracy, attributed to the enriched dataset provided by the GAN.

B. Studies Using Ensemble Learning Approaches

Ensemble learning approaches combine multiple models to improve classification accuracy and robustness. The following table provides an overview of studies that utilized ensemble learning techniques in skin disease classification.

Several studies have utilized ensemble learning approaches to enhance the classification of skin lesions. Popescu et al. [40] proposed an ensemble learning methodology that involved combining multiple models to classify skin lesions using the ISIC dataset. In this approach, different base models were trained separately, and their predictions were subsequently combined through majority voting or weighted averaging, leading to higher accuracy.

Similarly, Thurnhofer-Hemsi et al. [41] applied an ensemble approach to improve the accuracy of skin lesion classification on the HAM10000 dataset. This method focused on reducing the variance and bias inherent in single models by integrating the strengths of multiple classifiers, ultimately enhancing the overall classification performance.

C. Studies Using Multi-modal Approaches

Multi-modal techniques have been explored in several studies to enhance the accuracy of skin disease classification by integrating various data sources. Fu et al. [42] utilized a graph nodes-based approach that incorporated multi-modal data from different sources, including the 7-point checklist, ISIC2017, and ISIC2018 datasets. This methodology focused on combining different types of data to improve the classification of skin lesions.

Zhang et al. [43] introduced TFormer, a throughout fusion transformer model for multi-modal skin disease classification. The transformer model served as the feature extraction backbone, fusing both image and metadata, which significantly enhanced the accuracy of classification.

Roge et al. [28] employed a deep ensemble learning approach that integrated demographic information with image data to improve classification performance. This study combined the use of a demographic machine with standard imaging techniques, leading to superior results in classification accuracy.

Cai et al. [44] applied a multi-modal transformer model that fused images and metadata, leveraging the strengths of Vision Transformer (ViT) in handling multi-modal data. This approach provided a notable boost in classification accuracy by effectively combining different data sources.

Nguyen et al. [45] utilized a combination of deep learning models (DenseNet) and traditional machine learning algorithms (SVM) to classify skin lesions on imbalanced datasets. By incorporating both image data and patient metadata, this approach improved model robustness and accuracy.

Yin et al. [46] introduced the MetaNet module, which employs multi-modal data from both images and metadata to enhance skin tumor classification. The comprehensive analysis enabled by this multi-modal approach resulted in improved classification performance.

D. Studies Using Transformer-Based Models

Transformer-based models, originally developed for natural language processing, have been adapted for image classification tasks due to their ability to capture long-range dependencies in data.

Several studies have explored the use of Transformer-based models for skin lesion classification, showcasing the versatility and effectiveness of this approach. Ding et al. [47] employed a Vision Transformer (ViT) model on the ISIC2018, ISIC2017, and PH2 datasets. The methodology capitalized on the Transformer's ability to process image patches, which provided an effective solution for skin lesion classification.

Zhang et al. [43] proposed a Transformer-based model tailored for the Derm7pt dataset. Their approach emphasized the model's capability to learn global contextual information, which is crucial for accurately classifying skin lesions.

Abbas et al. [48] utilized a Transformer-based model to classify dermoscopic images from the ISIC dataset. The study highlighted the use of self-attention mechanisms inherent in Transformers, which allowed the model to capture intricate patterns in the images, thereby improving classification performance.

Cai et al. [44] applied a Vision Transformer (ViT) to the ISIC2018 dataset, focusing on the model's ability to process images in a patch-wise manner. Their study demonstrated the effectiveness of the Transformer architecture in handling large-scale image data, resulting in enhanced classification accuracy.

Yang et al. [49] also employed a Transformer-based model, specifically for the HAM10000 dataset. This study leveraged the self-attention mechanism of the Transformer to capture the intricate details of skin lesions, thereby improving classification accuracy.

Aladhadh et al. [50] used a Transformer-based model to classify images from the HAM10000 dataset. Their focus was on enhancing the model's sensitivity and specificity by taking advantage of the deep attention mechanisms inherent in Transformer architectures.

E. Studies Using Hybrid Models (CNN with Other Techniques)

Hybrid models that combine CNNs with other techniques, such as machine learning algorithms or other deep learning methods, have shown potential in improving classification accuracy.

Hybrid models that combine Convolutional Neural Networks (CNNs) with other techniques have been widely explored for improving the classification of skin lesions. Ma et al. [24] proposed a hybrid approach that integrated CNN with Random Forest for the classification of skin lesions using the HAM10000 dataset. In this study, the CNN was responsible for deep feature extraction, which were then classified by the Random Forest, combining the strengths of both methods.

Similarly, Roge et al. [28] developed a CNN-Random Forest hybrid model for the ISIC dataset (HAM10000). Their methodology also involved using CNN for initial feature extraction, followed by Random Forest for final classification, which enhanced both the accuracy and robustness of the model.

Nie et al. [51] explored a hybrid model that combined CNN with Transformer architecture, specifically for skin lesion classification using the ISIC2018 dataset. In this study, the CNN component handled initial feature extraction, while the Transformer captured global dependencies in the image data, leading to improved classification performance.

Li et al. [52] also utilized a hybrid CNN-SVM approach for the ISIC2019 dataset. In this study, the CNN was used for initial feature extraction, with SVM refining the classification process to enhance the accuracy of multi-class skin lesion classification.

Tahir et al. [29] proposed a hybrid model that combined CNN with XGBoost and Support Vector Machine (SVM) for classifying skin lesions in the ISIC2017 dataset. The CNN was utilized to extract relevant features, with XGBoost and SVM models performing the classification, leveraging the strengths of both deep learning and traditional machine learning techniques.

Khan et al [53] proposes a fully automated system for multiclass skin lesion localization and classification using deep learning. To address class imbalance in the HAM10000, ISBI2018, and ISBI2019 datasets, the approach fine-tunes a pre-trained DarkNet19 model and fuses visualized images using a High-Frequency approach with a Multilayered Feed-Forward Neural Network (HFaFFNN). Two additional models, DarkNet-53 and NasNet-Mobile, are trained using transfer learning on localized lesion images. Features are fused using the parallel max entropy correlation (PMEC) technique, and the entropy-kurtosis controlled whale optimization (EKWO) algorithm selects the most discriminant features. The model

achieves accuracies of 95.8%, 97.1%, and 85.35% on the HAM10000, ISBI2018, and ISBI2019 datasets, respectively. Zhou et Arandian. [54] introduce a new computer-aided skin cancer diagnosis method that combines deep learning with the Wildebeest Herd Optimization (WHO) algorithm. Inception CNN is used for feature extraction, followed by the WHO algorithm for feature selection to reduce analysis complexity. The method was tested on the ISIC-2008 dataset and compared with three other algorithms, demonstrating superior results.

Annaby et al. [55] proposes a melanoma detection approach that combines graph-theoretic representations with conventional image features to improve detection performance. Superpixels from dermoscopic images are used as graph nodes, and edges connect adjacent superpixels based on feature descriptor distances. Features are extracted from both weighted and unweighted graph models in the vertex and spectral domains, as well as from color, geometry, and texture. Various classifiers were trained on these feature combinations using ISIC datasets. The proposed system demonstrated significant improvements in accuracy, AUC, specificity, and sensitivity in detecting melanoma. Saeed et al. [56] employed a hybrid model that combined CNN with SVM for the classification of skin lesions using the ISIC2019 and ISIC2020 datasets. In this approach, the CNN handled feature extraction, while the SVM classifier provided enhanced decision boundaries, leading to improved model performance.

Afza et al. [57] introduced a hybrid CNN-Extreme Learning Machine (ELM) model for skin lesion classification using the ISIC dataset (HAM10000 and ISIC2018). In this study, CNN was used for feature extraction, while the ELM classifier processed these features, achieving high accuracy and specificity.

Nivedha et al. [58]focuses on melanoma detection using a novel computer-aided method combining the African Gorilla Troops Optimizer (AGTO) algorithm and Faster Region Convolutional Neural Networks (Faster R-CNN). The AGTO algorithm selects the most valuable features, and Faster R-CNN performs the classification. Applied to the ISIC-2020 skin cancer dataset, the proposed model achieves an accuracy of 98.55%, outperforming four existing approaches.

Damarla et al. [59] proposes an automated skin cancer classification system using a Deep Convolutional Neural Network (DCNN) for multiclass classification of dermoscopy images. The system employs transfer learning for feature extraction, followed by feature selection using metaheuristic algorithms like Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), and Gorilla Troop Optimization (GTO). A two-level classification is then applied to optimize and reduce features. Tested on the HAM10000 dataset, the model achieved 93.58% accuracy, outperforming state-of-the-art techniques and showing high scalability.

Finally, Attique Khan et al. [60] developed a two-stream deep neural network framework for multiclass skin cancer classification. The first stream uses a fusion-based contrast enhancement technique and DenseNet201 to extract and optimize features, while the second stream extracts features from a fine-tuned MobileNetV2. The most discriminant features from both streams are fused using a novel multimax correlation method and classified with an extreme learning machine. The classification of skin lesions was done using a mixed datasets (HAM10000, ISBI2018, ISIC2019).

F. Studies Using CNN-Based Models

Convolutional Neural Networks (CNNs) are a popular choice for image classification tasks due to their ability to automatically learn and extract features from images. The following table summarizes studies that employed CNN-based models for skin disease classification.

Numerous studies have explored the use of Convolutional Neural Networks (CNNs) for the classification of skin lesions, showcasing the versatility and effectiveness of these models. Alassaf et al. [25] utilized a deep learning-based Auto-Encoder model on the ISIC dataset, focusing on unsupervised feature learning. The model was trained to classify skin lesions based on these features, achieving high accuracy and specificity.

Abayomi et al. [61] proposes a novel data augmentation technique based on the covariant Synthetic Minority Oversampling Technique (SMOTE) to address data scarcity and class imbalance in melanoma detection. The augmented images, generated from the PH² dataset, are used to train the SqueezeNet deep learning model. In binary classification, the model achieved 92.18% accuracy, 80.77% sensitivity, and 95.1% specificity. In multiclass classification, it improved melanoma detection with 89.2% sensitivity and 96.2% specificity for atypical nevus detection, outperforming state-of-the-art methods.

Akram et al. [62] employed a CNN to classify dermoscopic images from the PH2, ISIC-MSK, and ISIC-UDA datasets. Their methodology centered on using CNN architectures for feature extraction and classification, resulting in high accuracy and sensitivity.

Calderon et al. [63] presents a bilinear CNN approach for classifying seven skin lesion types with high accuracy and low computational cost. The framework includes data augmentation to address class imbalance, transfer learning, and fine-tuning using ResNet50 and VGG16 architectures. Tested on the HAM10000 dataset, the model achieved a 2.7% improvement over the state-of-the-art.

Khan et al. [64] enhanced a deep learning architecture with the Entropy-NDOEM algorithm for multiclass skin lesion classification. The process used include contrast enhancement, fine-tuning EfficientNetB0 and DarkNet19, feature extraction and selection using Entropy-NDOELM, feature fusion, and classification using an extreme learning machine with diverse datasets such as HAM10000, ISIC2018, and ISIC2019 datasets. The overall methodology leads to an overall improvement in model performance.

Rasel et al. [65] presents in their study, a deep learning-based system using a Convolutional Neural Network (CNN) for melanoma detection to improve early diagnosis. The research focuses on how different nonlinear activation functions affect CNN performance on limited dermoscopic image datasets. The proposed model, using the parameterized Leaky ReLU function, achieved 97.5% accuracy, 98% precision, and 98% sensitivity in classifying skin lesions into three classes. Experiments were conducted on the PH2 and ISIC datasets,

demonstrating that this approach outperforms other activation functions for melanoma recognition.

Foahom et al. in their study [66] addresse the challenge of class imbalance in skin lesion (SL) detection by proposing an end-to-end decoupled training method for long-tailed skin lesion classification. The approach utilizes a novel loss function (Lf) for initial training to improve feature representation and a weighted variant of Lf to enhance robustness against class imbalance. Tested on the ISIC 2018 dataset, the model outperformed existing methods for SL detection by at least 2%, demonstrating its effectiveness in handling class imbalance.

Aldhyani et al. [26] proposes a lightweight deep learning model using dynamic-sized kernels for the accurate classification of skin lesions. The model incorporates both ReLU and leaky ReLU activation functions to enhance performance while maintaining a low number of trainable parameters. Tested on the HAM10000 dataset, the model achieved an impressive accuracy of 97.85%, outperforming several state-of-the-art models. The results demonstrate the model's efficiency in classifying various types of skin lesions.

Shen et al. [67] proposes a high-performance data augmentation strategy designed to improve skin cancer classification accuracy, particularly for use in low-resource settings. The strategy, which can be combined with any model in a plug-and-play mode, optimizes data augmentation with minimal computational cost. Using EfficientNets as a baseline, the model achieved strong performance on multiple datasets, including a BACC of 0.853 on HAM10000 and an AUC of 0.909 on ISIC 2017.

Kaur et al. [68] proposes an automated melanoma classifier using a deep convolutional neural network (DCNN) to classify malignant and benign melanoma from dermoscopic images. The DCNN is designed to efficiently extract features across multiple layers, optimizing filter selection, network depth, and hyperparameters to create a lightweight, less complex model. Tested on the ISIC 2016, 2017, and 2020 datasets, the model achieved accuracies of 81.41%, 88.23%, and 90.42%, respectively, outperforming other state-of-the-art methods and offering an efficient solution for melanoma diagnosis.

Dillshad et al. [69] focused on optimizing a CNN model for the classification of skin lesions in the HAM10000 dataset. The methodology based on a variance-controlled Marine Predator methodology optimizes feature selection and achieves high sensitivity and specificity in detecting various types of skin lesions.

Nugroho et al. [70] applied a CNN using the Inception-V3 architecture to the ISIC2019 dataset. Their approach emphasized feature extraction using the Inception-V3 model, followed by classification, which led to high accuracy in skin lesion detection. Mehmood et al. [71] proposed SBXception, a modified deep learning model based on the Xception architecture, designed for efficient skin lesion classification using the HAM10000 dataset. This methodology aim to enhance the original Xception model by making it shallower (reducing depth) and broader (increasing width), leading to fewer parameters and faster training times, achieving a high accuracy of 96.97% on the test set.

Mukadam et al. [72] used a CNN model to classify images from the HAM10000 dataset, with a focus on improving the CNN architecture to enhance accuracy and specificity in skin lesion classification.

Dahou et al. [73] proposes a robust skin cancer detection framework using a pre-trained MobileNetV3 architecture for feature extraction. The extracted features are optimized through a modified Hunger Games Search (HGS) algorithm, combining Particle Swarm Optimization and Dynamic-Opposite Learning (DOLHGS), to select the most relevant features. The model was tested on the ISIC-2016 and PH2 datasets and has improved classification accuracy across these datasets.

Finally, Supriyanto et al. [74] developed a CNN-based model for the classification of skin lesions using the HAM10000 dataset. The study focused on refining the CNN architecture to improve the sensitivity and specificity of the classification results.

G. Studies Using Machine Learning Techniques

Machine learning techniques such as XGBoost and SVM have been widely used for skin disease classification due to their robustness in handling complex datasets. The following table summarizes studies that employed these techniques, highlighting their methodologies.

Various studies have utilized machine learning techniques to improve the classification of skin lesions, often integrating multiple approaches to enhance performance. Khater et al. [75] employed XGBoost on selected features derived from dermoscopic images in the PH2 dataset. Their methodology focused on enhancing feature extraction through preprocessing and used explainable AI techniques, as SHAP to ensure the interpretability of the results, leading to improved classification accuracy.

Ahmed et al. [76] integrated Convolutional Neural Networks (CNNs) with Support Vector Machine (SVM) and Artificial Neural Network (ANN) models to analyze dermoscopic images from the HAM10000 and PH2 datasets. Their approach utilized MobileNet and ResNet101 architectures for feature extraction, followed by classification with SVM and ANN, achieving high accuracy across multiple metrics.

Ilkin et al. [77] proposed a combination of SVM with feature extraction from mixed datasets, including PH2 and ISIC. Their study focused on optimizing the feature set prior to SVM classification, aiming to improve the model's ability to differentiate between various skin lesion classes.

Finally, Pitchiah et al. [78] introduced a hybrid model combining K-Nearest Neighbors (KNN) with Random Forest and SVM. This model was validated on the PH2 dataset and aimed to balance sensitivity and precision by using ensemble methods to enhance classification performance.

These methodologies explanations serve as a focal point for the discussion. The detailed breakdown presented here assists researchers and practitioners in the identification of prevailing trends and gaps in the literature, and understanding which methodologies are most promising for clinical application and where further innovation may be required.

TABLE II. SUMMARY OF TECHNIQUES AND PERFORMANCE IN SYSTEMATIC REVIEW STUDIES

Ref.	Year	Technique	Data Type	Classes Number	Performance
[25]	2023	Deep learning-based Auto-Encoder	Image dataset (ISIC)	7	Accuracy: 96.83%, Sensitivity: 96.57%, Specificity: 97.83%
[76]	2023	CNN combined with SVM/ANN	HAM10000	7	HAM10000: Accuracy: 98.4%, Sensitivity: 94.46%, Specificity: 99.43%, AUC: 97.53%
			PH2	3	PH2 : Accuracy: 100%, Sensitivity: 100%, Specificity: 100%, AUC: 100%
[24]	2023	CNN integrated with Random Forest	HAM10000	7	Accuracy: 94.96%, Sensitivity: 93.74%, Specificity: 93.16%, F1-score: 93.24%
[75]	2023	XGBoost applied to selected features	PH2	3	Accuracy: 94%, AUC: 99.47%
[64]	2023	CNN enhanced with ELM (Extreme Learning Machine)	HAM10000	7	Accuracy : 95.7%
			ISIC2018	7	Accuracy : 96.3%
			ISIC2019	8	Accuracy : 94.8%
[47]	2023	Vision Transformer (ViT) model	ISIC2018	8	Accuracy: 93.2%, Specificity: 92.2%, AUC: 97.7%
			ISIC2017	2	Accuracy: 89.5%, Specificity: 94.7%, AUC: 96.2%
			PH2	2	Accuracy: 91.4%, Specificity: 88.5%, AUC: 96.3
[62]	2023	Deep models with entropy-controlled optimization for feature selection	PH2	3	Accuracy: 98.89%, Specificity: 98.9%, Sensivity: 98%
			ISIC-MSK	Various	Accuracy: 99.01%, Specificity: 99.4%, Sensivity: 98.5%
			ISIC-UDA	3	Accuracy: 99.09%, Specificity: 99.4%, Sensivity: 98.6%
[43]	2023	Transformer-based model	Image dataset (Derm7pt)	2	Accuracy: 80.03%
[51]	2023	Hybrid CNN and Transformer model	Image dataset (ISIC2018)	7	F1 score: 87.37%, Sensitivity: 88.13%, Specificity: 88.29%
[29]	2023	DSCC_Net model with SMOTE Tomek	ISIC2020, Derm-IS, HAM10000	Various	Accuracy: 94.17%, Sensitivity: 94.28%, Specificity: 93.76%, F1-score: 93.93%
[58]	2023	Region-based CNN (RCNN)	Image dataset (ISIC2020)	2	Accuracy: 98.55%, Sensitivity: 96.92%, Specificity: 98.11%, Precision: 98.34%
[48]	2023	Transformer-based model	Image dataset (Personalized ISIC)	9	Accuracy: 95.6%, Sensitivity: 96.7%, Specificity: 95%
[69]	2023	Optimization-aided deep learning with MobileNetV2, NasNet, and Marine Predator	Image dataset (HAM10000)	7	Accuracy: 94.4%, Sensitivity: 94.4%, Specificity: 94.4%
[70]	2023	Convolutional Neural Network (CNN) using Inception-V3	ISIC dataset (ISIC2019)	7	Accuracy: 96.4%, AUC: 0.98
[28]	2023	Hybrid CNN with Random Forest	HAM10000	7	AUC: 87.6%
[72]	2023	Custom CNN with ESRGAN preprocessing	HAM10000	7	Accuracy: 98.9%
[73]	2023	CNN with Hunger Games Search	PH2	3	Accuracy: 96.43%
			ISIC2016	2	Accuracy: 88.19%
[44]	2023	Multi-modal Vision Transformer (ViT) model	ISIC dataset (ISIC2018)	7	Accuracy: 93.81%, Sensitivity: 90.14%, Specificity: 98.36%, F1-score: 90.13 %

Ref.	Year	Technique	Data Type	Classes Number	Performance
[49]	2023	Transformer-based model	HAM10000	7	Accuracy: 94.1%
[56]	2023	Hybrid CNN with SVM	ISIC2019	8	Accuracy: 96%
			ISIC2020	2	Accuracy: 92%
[71]	2023	SbXception: an enhanced Xception architecture	HAM10000	7	Accuracy: 96.97%, Sensitivity: 95.43%, Specificity: 85.34%
[74]	2023	Two-stage image augmentation with GAN and CNN models	HAM10000	7	Accuracy: 96.9%, Sensitivity: 96.87%, Specificity: 97.07%, F1-score: 96.97%
[52]	2022	Hybrid CNN with SVM	ISIC dataset (ISIC2019)	8	Accuracy: 98.76%, Sensitivity: 98.4%, Specificity: 99.81%
[59]	2022	A deep learning system using transfer learning and meta-heuristic optimization	HAM10000	7	Accuracy: 93.58%
[67]	2022	Convolutional Neural Network (CNN) using EfficientNets	HAM10000	7	Accuracy: 95.8%, Sensitivity: 85.3%, Specificity: 97.9%, AUC: 0.975
[45]	2022	Convolutional Neural Network with Soft-attention	HAM10000	7	Accuracy: 90%, Sensitivity: 81%, FA-score: 81%, AUC: 0.99
[63]	2021	Convolutional Neural Network (CNN) using bilinear approach	HAM10000	7	Accuracy: 93.21%, Sensitivity: 93%, Specificity: 92.92% , AUC: 0.98
[61]	2021	Convolutional Neural Network using SMOTE Oversampling Technique	Image dataset (PH2)	3	Accuracy: 92.18%, Sensitivity: 80.77%, Specificity: 95.1%, F1-score: 80.84%
[55]	2021	Graph nodes in CNN architecture	ISIC datasets	7	Accuracy: 97.4%, Sensitivity: 100%, Specificity: 95.16%, AUC: 99.91%
[68]	2022	Deep Convolutional Neural Network (DCNN)	ISIC2016	2	Accuracy: 81.41%
			ISIC2017	2	Accuracy: 88.23%
			ISIC2020	2	Accuracy: 90.42%
[65]	2022	Convolutional Neural Network (CNN) Leaky ReLU function	Mixed dataset (PH2, ISIC)	2,7	Accuracy: 97.5%, Precision: 98.0%, Sensitivity: 98.0%
[77]	2021	Support Vector Machine (SVM)	PH2	3	Accuracy: 97.5%, Sensitivity: 93.75%, Specificity: 100%, AUC: 97%
			ISIC	3	Accuracy: 97.56%, Sensitivity: 97.94%, Specificity: 97.07%, AUC: 98%
[57]	2022	Hybrid CNN-ELM (Extreme Learning Machine)	HAM1000	7	Accuracy: 93.4%, Precision: 93.10%
			ISIC2018	7	Accuracy: 94.36%, Precision: 94.08%
[46]	2022	Hybrid CNN with DenseNet-169	Mixed dataset (PAD-UFES-20, ISIC 2019)	7	Accuracy: 81.4%
[66]	2022	Convolutional Neural Network (CNN)	ISIC dataset (ISIC2018)	7	B.Accuracy: 88%
[54]	2021	CNN combined with Meta-Heuristic methods	ISIC dataset (ISIC2008)	Not defined	Accuracy: 96%, Sensitivity: 96%, Specificity: 95%
[26]	2022	Lightweight CNN with dynamic-sized kernels and ReLU/leaky ReLU activations	HAM10000	7	Accuracy: 97.8%, Sensitivity: 98%, Specificity: 98.1%, F1-score: 98%
[39]	2021	GAN combined with CNN	ISIC dataset (ISIC2018)	7	Accuracy: 70.1%
[78]	2022	K-Nearest Neighbors (KNN) combined with Random Forest (RF) and SVM	Image dataset (PH2)	3	Accuracy: 94.81%

Ref.	Year	Technique	Data Type	Classes Number	Performance
[53]	2021	CNN with PMEC feature fusion and EKWO optimization	HAM10000	7	Accuracy: 95.8%
			ISIC2018	7	Accuracy: 97.1%
			ISBI2019	8	Accuracy: 85.35%
[79]	2022	Convolutional Neural Network (CNN)	HAM10000	7	Accuracy: 91%, F1-score: 88.1%, ROC-AUC: 95%
[50]	2022	Transformer-based model	HAM10000	7	Accuracy: 96.14%, Sensitivity: 96.5%, Specificity: 96%, F1-score:97%
[80]	2021	CNN combined with Machine Learning (ML) techniques	HAM10000	7	Accuracy: 91.7%
[30]	2022	Explainable AI (XAI) approach using CNN	ISIC dataset (ISIC2019)	8	Accuracy: 94.47%, Sensitivity: 94.01%, Specificity: 93.57%, F1-score: 94.45%
[81]	2021	Neural Network (NN)	ISIC dataset (ISIC2017)	2	Precision: 94%, Sensitivity: 93%, Specificity: 91%
[82]	2021	Feedforward Neural Network combined with Artificial Neural Network	ISIC2018	7	Accuracy: 90%, Sensitivity: 89.37%, Specificity: 97.84%
			PH2	3	Accuracy: 95.8%, Sensitivity: 95.64%, Specificity: 98.21%
[83]	2023	CLCM-net model with layer-wise weight constraints	ISIC2018	7	Accuracy: 94.42%
			ISIC2019	8	Accuracy: 95.8%
			Combined dataset	-	Accuracy: 93%
[84]	2022	Capsule Network	HAM10000	7	Accuracy: 96.49%
[85]	2021	Neural Network (NN)	7-Point	2	Accuracy: 95.42%, Sensitivity: 98.01%, Specificity: 94.4%
			Med-Node	2	Accuracy: 94.71%, Sensitivity: 96.42%, Specificity: 87.5%
			PH2	3	Accuracy: 94.88%, Sensitivity: 100%, Specificity: 85.62%
[60]	2022	Two-stream neural network with feature fusion and ELM	HAM10000	7	Accuracy: 96.5%
			ISBI2018	7	Accuracy: 98%
			ISIC2019	8	Accuracy: 89%
[40]	2022	Ensemble Learning approach	HAM10000	7	Validation Accuracy: 86.71%
[86]	2022	StochasticProgressive Instance Learning	ISIC2017	2	Accuracy: 88%, AUC: 98.3
			ISIC2018	7	Accuracy: 89.4%, AUC: 92.9
[42]	2022	Graph nodes-based approach	7-point	2	AUC: 83.6%
[87]	2021	CNN combined with ANN	ISIC dataset	5	Balanced Accuracy: 92.34%, Sensitivity: 87.10%, Specificity: 94.19% , AUC-ROC: 97.10%
[41]	2021	Ensemble Learning approach	HAM10000	7	Accuracy: 83.5%, Sensitivity: 65.6%, Specificity: 95.4%
[27]	2024	Generative Adversarial Network (GAN) with CNN	HAM10000	7	Accuracy: 98.23%, Sensitivity: 88.85%, Specificity: 98.34%, F1-score:89.48

The varying performance of AI models across different datasets can be attributed to the unique characteristics of each dataset, such as size, number of classes, image diversity, and data quality. Larger datasets like ISIC2018 and ISIC2019, which contain thousands of high-resolution dermoscopic images, generally allow AI models to perform well due to the availability of diverse training data. However, the number of classes also plays a crucial role. For example, studies using ISIC2019, which contains 8 classes, reported lower accuracy (e.g. [64] achieved 94.8% accuracy) compared to ISIC2018 (96.3% accuracy), which has fewer classes and is less complex. The increased difficulty in distinguishing between more skin disease types likely accounts for this discrepancy. Conversely, smaller datasets like PH2, with only 200 images and 3 classes, often result in higher accuracy within the dataset, as seen in [62] with an accuracy of 98.89%. However, such models may struggle with generalizability when applied to larger, more complex datasets. Similarly, HAM10000 provides a diverse dataset with 7 classes, leading to good overall performance in most models (e.g., 95.7% accuracy in [64]), although class imbalance in some categories can affect results. Models tested on clinical image datasets like Derm7pt, which include more background noise and variability in image conditions, often report lower performance compared to dermoscopic datasets, as these models must handle more variability in input data. This analysis suggests that certain algorithms, particularly CNNs and hybrid models, are better suited to high-quality, diverse dermoscopic datasets, while challenges like class imbalance, dataset size, and image quality play a significant role in the comparative results across different datasets.

The systematic review of recent studies on AI models for skin disease classification, as detailed in Table II and categorized in the paragraphs above, reveals several critical trends, patterns, and challenges that underscore the current state of research in this domain. To provide a nuanced understanding of these approaches, Fig. 8 presents the distribution of the AI techniques utilized across the studies. This figure highlights the prevalence of different methodologies, offering insights into the current state of AI-driven skin disease classification.

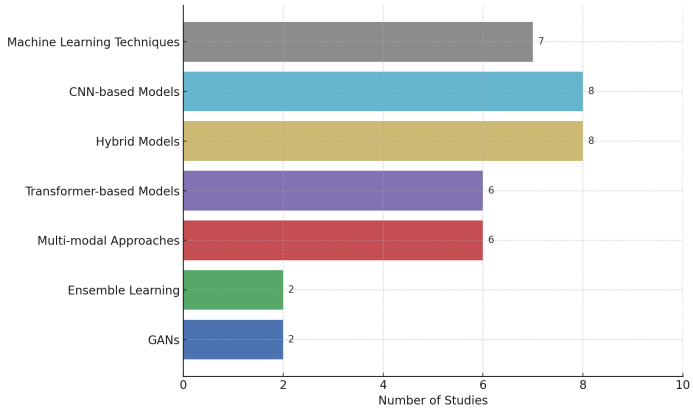


Fig. 8. Trends in AI techniques for skin disease classification (2021-2023).

A significant trend observed across the reviewed literature is the dominance of Convolutional Neural Networks (CNNs) and their variations, including hybrid models combining CNNs with other machine learning techniques such as Support Vector Machines (SVMs), Extreme Learning Machines (ELMs), and Random Forests. This reflects the strong performance of CNNs in image recognition tasks, which are central to dermatological diagnosis. To further analyze the methodologies within CNN-based models and their variations, Fig. 9 presents the usage frequency of various pre-trained CNN models across the studies. This figure provides a clear overview of which pre-trained models are most commonly employed in skin disease classification.

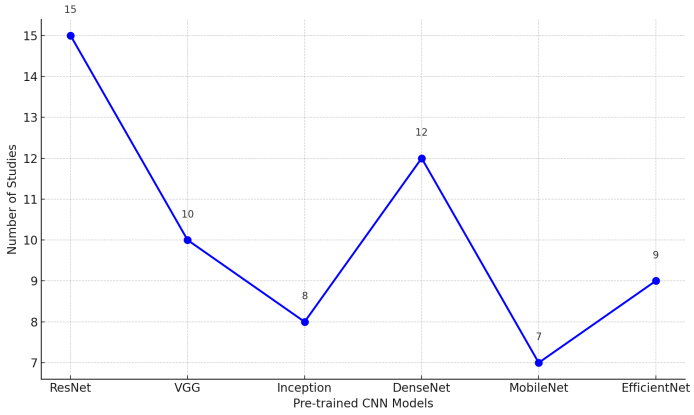


Fig. 9. Usage frequency of Pre-trained CNN models in skin disease classification studies.

More recently, the emergence of Transformer-based models and Vision Transformers (ViTs) suggests a growing interest in leveraging these advanced architectures, which have demonstrated exceptional performance in natural language processing and are now being adapted for medical imaging.

The reviewed studies span from 2021 to 2023, indicating a rapid evolution of techniques over a relatively short period. Early studies primarily focused on straightforward CNN architectures, while later studies have increasingly explored more complex hybrid models and the integration of transformers. This evolution suggests a shift towards more sophisticated, multi-faceted approaches aimed at improving model accuracy and generalization across diverse datasets. Fig. 10 illustrates the increasing trend in the use of Hybrid Models and Transformers over time.

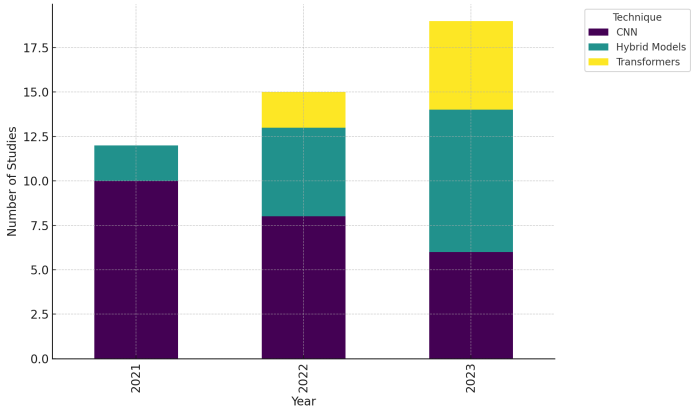


Fig. 10. Trends in AI techniques for skin disease classification (2021-2023).

Additionally, there is a noticeable increase in the application of Auto-Encoders, Generative Adversarial Networks (GANs), and ensemble methods, reflecting the ongoing effort to enhance model robustness and address challenges such as data scarcity and imbalance. The development of GANs combined with CNNs, as seen in recent studies, highlights a trend towards generating synthetic data to improve training outcomes, which is particularly valuable in medical fields where annotated data can be limited.

The predominant use of datasets such as HAM10000, ISIC (International Skin Imaging Collaboration), and PH2 across the studies indicates a reliance on these well-established, publicly available image datasets. The consistent use of these datasets underscores their role as benchmarks in the field. However, it also highlights a potential limitation in terms of dataset diversity, as most studies focus on the same data sources, which may not fully represent the variety of skin conditions encountered in clinical practice.

In terms of class distribution, most studies focus on classifying a limited number of conditions, with a particular emphasis on melanoma and non-melanoma skin cancers. This trend reflects the clinical importance of accurately diagnosing these conditions but also points to a gap in research focused on rarer or less visually distinct skin diseases, which are underrepresented in current models.

The reported performance metrics, including accuracy, sensitivity, specificity, and precision, show high variability across studies, with accuracy ranging from 88.19% to 98.7%. This variability can be attributed to differences in model complexity, data preprocessing methods, and the inherent difficulty of the classification tasks. Notably, hybrid models and those incorporating transformers generally report higher performance, suggesting that these more complex models may offer advantages in handling the nuances of skin disease classification. Fig. 11 provides a comparative analysis of the performance distribution across different model types.

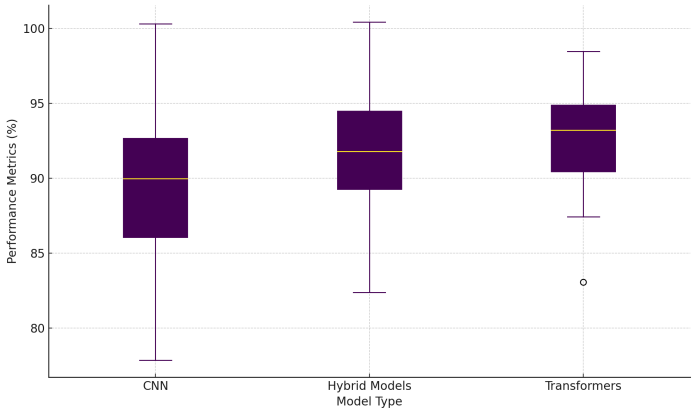


Fig. 11. Performance distribution across different models.

However, it is important to note that high performance on well-curated datasets does not necessarily translate to clinical effectiveness. The consistent reporting of high specificity across models is encouraging, as it suggests a strong ability to correctly identify negative cases, which is critical in avoiding unnecessary interventions. Nonetheless, the lower sensitivity scores reported in some studies, particularly those involving simpler CNN architectures, indicate a potential risk of missed diagnoses, which could have serious clinical implications.

Despite the progress made, several challenges remain. The reliance on a few datasets raises concerns about the generalizability of these models to broader, more diverse patient populations. Additionally, while the integration of multimodal data (e.g., combining images with patient history) is increasingly being recognized as essential for improving diagnostic accuracy, few studies have fully implemented this approach.

Moreover, the computational complexity of advanced models such as transformers and GANs may limit their deployment in resource-constrained settings, highlighting the need for research focused on optimizing these models for real-world clinical environments. Fig. 12 compares the performance metrics of hybrid models versus pure deep learning models, emphasizing the potential advantages of hybrid approaches.

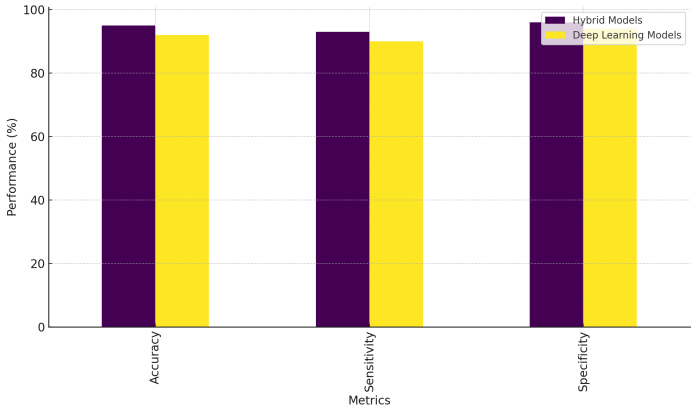


Fig. 12. Performance comparison: Hybrid models vs pure deep learning.

The general metrics commonly used to evaluate the performance of AI models in skin disease classification include Accuracy, Sensitivity, Specificity, Precision, F1-Score, and Area Under the Curve (AUC). These metrics are crucial in assessing the effectiveness of the models in correctly identifying and classifying skin conditions.

1. Accuracy: The proportion of correctly classified instances (both true positives and true negatives) among the total number of instances.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Sensitivity (Recall or True Positive Rate): The proportion of actual positives that are correctly identified by the model.

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

3. Specificity (True Negative Rate): The proportion of actual negatives that are correctly identified by the model.

$$\text{Specificity} = \frac{TN}{TN + FP}$$

4. Precision (Positive Predictive Value): The proportion of positive predictions that are actually correct.

$$\text{Precision} = \frac{TP}{TP + FP}$$

5. F1-Score: The harmonic mean of Precision and Sensitivity, providing a balance between the two.

$$F1\text{-Score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}$$

6. Area Under the Curve (AUC): The area under the Receiver Operating Characteristic (ROC) curve, representing the model's ability to discriminate between classes.

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}^{-1}(x))\, dx$$

VII. Conclusion

This systematic review provides a focused analysis of AI methodologies in skin disease classification, highlighting the growing adoption of advanced techniques like GANs, Transformer models, and multi-modal approaches. While CNNs remain a dominant tool, their performance is often enhanced by hybrid and ensemble learning methods, demonstrating a trend towards more complex model architectures.

However, this review also identifies several key challenges that need to be addressed. The lack of standardization across studies and the limited application of multi-modal approaches restrict the generalizability of current models. Additionally, the datasets used in most studies lack diversity in terms of patient demographics, including under-representation of different skin tones and rare skin diseases, which limits the models' applicability across broader populations. Moreover, the explainability of AI models remains a critical barrier to their integration into clinical practice. Ensuring that AI-driven diagnostic tools provide transparent and interpretable outputs for clinicians is crucial for their adoption in real-world settings.

Future research should prioritize the development of standardized protocols and benchmarking methods to enable meaningful comparisons between different AI models. The creation of more diverse and representative datasets is essential to improve the generalization of AI models and their applicability in real-world clinical environments. Furthermore, addressing the ethical implications of AI use in dermatology is vital, particularly in relation to bias mitigation, ensuring that AI technologies perform equitably across all patient groups. There is also a growing need to investigate advanced techniques such as federated learning, which could enhance collaboration between institutions while preserving patient privacy, thereby improving model generalizability without compromising data security.

Finally, real-world validation through clinical trials and large-scale implementation studies is necessary to evaluate the practical utility and reliability of AI-driven skin disease classification tools. Collaboration between AI researchers and clinicians will be crucial in translating these models from research into clinical practice.

References

[1] H. Kutzner, T. B. Jutzi, D. Krahl, E. I. Krieghoff-Henning, M. V. Heppt, A. Hekler, M. Schmitt, R. C. Maron, S. Fröhling, C. von Kalle *et al.*, "Overdiagnosis of melanoma–causes, consequences and solutions," *JDDG: Journal der Deutschen Dermatologischen Gesellschaft*, vol. 18, no. 11, pp. 1236–1243, 2020.

[2] A. A. Marghoob, L. Changchien, J. DeFazio, W. C. Dessio, J. Malvehy, I. Zalaudek, A. C. Halpern, and A. Scope, "The most common challenges in melanoma diagnosis and how to avoid them," *Australasian Journal of Dermatology*, vol. 50, no. 1, pp. 1–13, 2009.

[3] C. Karimkhani, A. C. Green, T. Nijsten, M. A. Weinstock, R. P. Dellavalle, M. Naghavi, and C. Fitzmaurice, "The global burden of melanoma: results from the global burden of disease study 2015," *British Journal of Dermatology*, vol. 177, no. 1, pp. 134–140, 2017.

[4] C. E. Griffiths, A. W. Armstrong, J. E. Gudjonsson, and J. N. Barker, "Psoriasis," *The Lancet*, vol. 397, no. 10281, pp. 1301–1315, 2021.

[5] W. Peng and N. Novak, "Pathogenesis of atopic dermatitis," *Clinical & Experimental Allergy*, vol. 45, no. 3, pp. 566–574, 2015.

[6] L. Thomas and S. Puig, "Dermoscopy, digital dermoscopy and other diagnostic tools in the early detection of melanoma and follow-up of high-risk skin cancer patients." *Acta dermato-venereologica*, vol. 97, 2017.

[7] R. A. Shellenberger, F. Fayyaz, Z. Sako, M. Schaeffer, K. Tawagi, C. Scheidel, and M. Nabhan, "Impact of biopsy technique on clinically important outcomes for cutaneous melanoma: a systematic review and meta-analysis," *Mayo Clinic Proceedings: Innovations, Quality & Outcomes*, vol. 4, no. 4, pp. 373–383, 2020.

[8] P. Carli, D. Massi, V. De Giorgi, and B. Giannotti, "Reliability and inter-observer agreement of dermoscopic diagnosis of melanocytic lesions," *Journal of the American Academy of Dermatology*, vol. 48, no. 6, pp. 878–883, 2003.

[9] D. A. Gutman, N. C. Codella, M. E. Celebi, B. Helba, M. A. Marchetti, N. Mishra, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016 hosted by the international skin imaging collaboration (isic)," *arXiv preprint arXiv:1605.01397*, 2016.

[10] N. Noroozi and A. Zakerolhosseini, "Differential diagnosis of squamous cell carcinoma in situ using skin histopathological images," *Computers in biology and medicine*, vol. 70, pp. 23–39, 2016.

[11] H. D. Heibel, L. Hooey, and C. J. Cockerell, "A review of noninvasive techniques for skin cancer detection in dermatology," *American journal of clinical dermatology*, vol. 21, no. 4, pp. 513–524, 2020.

[12] S. L. Schneider, I. Kohli, I. H. Hamzavi, M. L. Council, A. M. Rossi, and D. M. Ozog, "Emerging imaging technologies in dermatology: Part ii: Applications and limitations," *Journal of the American Academy of Dermatology*, vol. 80, no. 4, pp. 1121–1131, 2019.

[13] S. R. Jartarkar, A. Patil, U. Wollina, M. H. Gold, H. Stege, S. Grabbe, and M. Goldust, "New diagnostic and imaging technologies in dermatology," *Journal of cosmetic dermatology*, vol. 20, no. 12, pp. 3782–3787, 2021.

[14] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, 2017.

[15] M. Goyal, T. Knackstedt, S. Yan, and S. Hassanpour, "Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities," *Computers in biology and medicine*, vol. 127, p. 104065, 2020.

[16] M. A. Thaajwer and U. P. Ishanka, "Melanoma skin cancer detection using image processing and machine learning techniques," in *2020 2nd international conference on advancements in computing (ICAC)*, vol. 1. IEEE, 2020, pp. 363–368.

[17] T. Shanthi, R. Sabeenian, and R. Anand, "Automatic diagnosis of skin diseases using convolution neural network," *Microprocessors and Microsystems*, vol. 76, p. 103074, 2020.

[18] T. J. Brinker, A. Hekler, A. H. Enk, C. Berking, S. Haferkamp, A. Hauschild, K. Ghoreschi, M. Weichenthal, J. Klode, D. Schadendorf *et al.*, "Skin cancer classification using convolutional neural networks: systematic review," *Journal of the American Academy of Dermatology*, vol. 79, no. 6, pp. 1233–1242, 2018.

[19] J. Zhang, F. Zhong, K. He, M. Ji, S. Li, and C. Li, "Recent advancements and perspectives in the diagnosis of skin diseases using machine learning and deep learning: A review," *Diagnostics*, vol. 13, no. 23, p. 3506, 2023.

[20] X. Liu, L. Faes, A. U. Kale, S. K. Wagner, D. J. Fu, A. Bruynseels, T. Mahendiran, G. Moraes, M. Shamdas, C. Kern *et al.*, "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," *The lancet digital health*, vol. 1, no. 6, pp. e271–e297, 2019.

[21] P. Tschandl, C. Rinner, Z. Apalla, G. Argenziano, N. Codella, A. Halpern, J. Malvehy, S. Puig, C. Rosendahl, and H. Kittler, "Human-computer collaboration for skin cancer recognition," *Nature Medicine*, vol. 26, no. 8, pp. 1229–1234, 2020.

[22] Y. Wu, B. Chen, A. Zeng, D. Pan, R. Wang, and S. Zhao, "Skin cancer classification with deep learning: a systematic review," *Frontiers in Oncology*, vol. 12, p. 893972, 2022.

[23] E. R. Gordon, M. H. Trager, D. Kontos, C. Weng, L. J. Geskin, L. S. Dugdale, and F. H. Samie, "Ethical considerations for artificial intelligence in dermatology: a scoping review," *British Journal of Dermatology*, p. ljae040, 2024.

[24] X. Ma, J. Shan, F. Ning, W. Li, and H. Li, "Effnet: A skin cancer classification model based on feature fusion and random forests," *PLOS ONE*, vol. 18, no. 10, pp. 1–17, 10 2023.

[25] A. Alassaf, "Deep learning-based stacked auto-encoder with dynamic differential annealed optimization for skin lesion diagnosis," *Computer Systems Science and Engineering*, vol. 47, no. 3, pp. 2773–2789, 2023.

[26] T. H. Aldhyani, A. Verma, M. H. Al-Adhaileh, and D. Koundal, "Multi-class skin lesion classification using a lightweight dynamic kernel deep-learning-based convolutional neural network," *Diagnostics*, vol. 12, no. 9, p. 2048, 2022.

[27] Q. Su, H. N. A. Hamed, M. A. Isa, X. Hao, and X. Dai, "A gan-based data augmentation method for imbalanced multi-class skin lesion classification," *IEEE Access*, 2024.

[28] A. Roge, P. Ting, A. Chern, and W. Ting, "Deep ensemble learning using a demographic machine learning risk stratifier for binary classification of skin lesions using dermatoscopic images," *Journal of Medical Artificial Intelligence*, vol. 6, no. 0, 2023.

[29] M. Tahir, A. Naeem, H. Malik, J. Tanveer, R. A. Naqvi, and S.-W. Lee, "Dscc_net: Multiclassification deep learning models for diagnosing of skin cancer using dermoscopic images," *Cancers*, vol. 15, no. 7, 2023.

[30] N. Nigar, M. Umar, M. K. Shahzad, S. Islam, and S. Abawi, "A deep learning approach based on explainable ai for skin lesion classification," *IEEE Access*, 2022.

[31] B. Cassidy, C. Kendrick, A. Brodzicki, J. Jaworek-Korjakowska, and M. H. Yap, "Analysis of the isic image datasets: Usage, benchmarks and recommendations," *Medical image analysis*, vol. 75, p. 102305, 2022.

[32] P. Tschandl, C. Rosendahl, and H. Kittler, "The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific data*, vol. 5, p. 180161, 2018.

[33] T. Mendonça, M. Celebi, T. Mendonca, and J. Marques, "Ph2: A public database for the analysis of dermoscopic images," *Dermoscopy image analysis*, vol. 2, 2015.

[34] J. Kawahara, S. Daneshvar, G. Argenziano, and G. Hamarneh, "Seven-point checklist and skin lesion classification using multitask multimodal neural nets," *IEEE journal of biomedical and health informatics*, vol. 23, no. 2, pp. 538–546, 2018.

[35] A. G. Pacheco, G. R. Lima, A. S. Salomao, B. Krohling, I. P. Biral, G. G. de Angelo, F. C. Alves Jr, J. G. Esgario, A. C. Simora, P. B. Castro *et al.*, "Pad-ufes-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones," *Data in brief*, vol. 32, p. 106221, 2020.

[36] B. Zhang, X. Zhou, Y. Luo, H. Zhang, H. Yang, J. Ma, and L. Ma, "Opportunities and challenges: Classification of skin disease based on deep learning," *Chinese Journal of Mechanical Engineering*, vol. 34, pp. 1–14, 2021.

[37] A. Cascarano, J. Mur-Petit, J. Hernandez-Gonzalez, M. Camacho, N. de Toro Eadie, P. Gkontra, M. Chadeau-Hyam, J. Vitria, and K. Lekadir, "Machine and deep learning for longitudinal biomedical data: a review of methods and applications," *Artificial Intelligence Review*, vol. 56, no. Suppl 2, pp. 1711–1771, 2023.

[38] K. Frasier, V. Li, M. Coleman, E. Rodriguez, T. Karatas *et al.*, "Digital intelligence for predicting skin disease progression and treatment outcomes," *Ameri J Clin Med Re: AJCMR-137*, 2024.

[39] I. S. A. Abdelhalim, M. F. Mohamed, and Y. B. Mahdy, "Data augmentation for skin lesion using self-attention based progressive generative adversarial network," *Expert Systems with Applications*, vol. 165, p. 113922, 2021.

[40] D. Popescu, M. El-Khatib, and L. Ichim, "Skin lesion classification using collective intelligence and deep learning," *Sensors*, 2022.

[41] K. Thurnhofer-Hemsi, E. López-Rubio, E. Dominguez, and D. A. Elizondo, "Skin lesion classification by ensembles of deep convolutional networks and regularly spaced shifting," *IEEE Access*, vol. 9, pp. 112 193–112 205, 2021.

[42] X. Fu, L. Bi, A. Kumar, M. Fulham, and J. Kim, "Graph-based intercategory and intermodality network for multilabel classification and melanoma diagnosis of skin lesions in dermoscopy and clinical images," *IEEE Transactions on Medical Imaging*, vol. 41, no. 11, pp. 3266–3277, 2022.

[43] Y. Zhang, F. Xie, and J. Chen, "Tformer: A throughout fusion transformer for multi-modal skin lesion diagnosis," *Computers in Biology and Medicine*, vol. 157, p. 106712, 2023.

[44] G. Cai, Y. Zhu, Y. Wu, X. Jiang, J. Ye, and D. Yang, "A multimodal transformer to fuse images and metadata for skin disease classification," *The Visual Computer*, vol. 39, no. 7, pp. 2781–2793, 2023.

[45] V. D. Nguyen, N. D. Bui, and H. K. Do, "Skin lesion classification on imbalanced data using deep learning with soft attention," *Sensors*, vol. 22, no. 19, p. 7530, 2022.

[46] W. Yin, J. Huang, J. Chen, and Y. Ji, "A study on skin tumor classification based on dense convolutional networks with fused metadata," *Frontiers in Oncology*, vol. 12, p. 989894, 2022.

[47] Y. Ding, Z. Yi, M. Li, J. long, S. Lei, Y. Guo, P. Fan, C. Zuo, and Y. Wang, "Hi-mvit: A lightweight model for explainable skin disease classification based on modified mobilevit," *Digital Health*, vol. 9, p. 20552076231207197, 2023.

[48] Q. Abbas, Y. Daadaa, U. Rashid, and M. E. Ibrahim, "Assist-dermo: A lightweight separable vision transformer model for multiclass skin lesion classification," *Diagnostics*, vol. 13, no. 15, p. 2531, 2023.

[49] G. Yang, S. Luo, and P. Greer, "A novel vision transformer model for skin cancer detection," *Neural Processing Letters*, 2023.

[50] S. Aladhadh, M. Alsanea, M. Aloraini, T. Khan, and U. Tariq, "An effective skin cancer classification mechanism using hybrid cnn and svm models," *Sensors*, 2022.

[51] Y. Nie, P. Sommella, M. Carratù, M. O'Nils, and J. Lundgren, "A deep cnn transformer hybrid model for skin lesion classification of dermoscopic images using focal loss," *Diagnostics*, vol. 13, no. 1, 2023.

[52] Z. Li, Z. Chen, X. Che, Y. Wu, D. Huang, H. Ma, and Y. Dong, "A classification method for multi-class skin damage images combining quantum computing and inception-resnet-v1," *Frontiers in Physics*, vol. 10, p. 1046314, 2022.

[53] M. A. Khan, T. Akram, M. Sharif, S. Kadry, and Y. Nam, "Computer decision support system for skin cancer localization and classification." *Computers, Materials & Continua*, vol. 68, no. 1, 2021.

[54] B. Zhou and B. Arandian, "An improved cnn architecture to diagnose skin cancer in dermoscopic images based on wildebeest herd optimization algorithm," *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, p. 7567870, 2021.

[55] M. H. Annaby, A. M. Elwer, M. A. Rushdi, and M. E. Rasmy, "Melanoma detection using spatial and spectral analysis on superpixel graphs," *Journal of digital imaging*, vol. 34, no. 1, pp. 162–181, 2021.

[56] M. Saeed, A. Naseer, H. Masood, S. U. Rehman, and V. Gruhn, "The power of generative ai to augment for enhanced skin cancer classification: A deep learning approach," *IEEE Access*, vol. 11, pp. 130 330–130 344, 2023.

[57] F. Afza, M. Sharif, M. A. Khan, U. Tariq, H.-S. Yong, and J. Cha, "Multiclass skin lesion classification using hybrid deep features selection and extreme learning machine," *Sensors*, vol. 22, no. 3, p. 799, 2022.

[58] S. Nivedha and S. Shankar, "Melanoma diagnosis using enhanced faster region convolutional neural networks optimized by artificial gorilla troops algorithm," *Information Technology and Control*, vol. 52, no. 4, pp. 819–832, 2023.

[59] A. Damarla and D. Sumathi, "An approach for optimization of features using gorilla troop optimizer for classification of melanoma," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, 2022.

[60] M. Attique Khan, M. Sharif, T. Akram, S. Kadry, and C.-H. Hsu, "A two-stream deep neural network-based intelligent system for complex skin cancer types classification," *International Journal of Intelligent Systems*, vol. 37, no. 12, pp. 10 621–10 649, 2022.

[61] O. O. Abayomi-Alli, R. Damasevicius, S. Misra, R. Maskeliunas, and A. Abayomi-Alli, "Malignant skin melanoma detection using image augmentation by oversamplingin nonlinear lower-dimensional embedding manifold," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 29, no. 8, pp. 2600–2614, 2021.

[62] T. Akram, R. Junejo, A. Alsuhaibani, M. Rafiullah, A. Akram, and N. A. Almujally, "Precision in dermatology: Developing an optimal feature selection framework for skin lesion classification," *Diagnostics*, vol. 13, no. 17, 2023.

[63] C. Calderón, K. Sanchez, S. Castillo, and H. Arguello, "Bilsk: A bilinear convolutional neural network approach for skin lesion classification," *Computer Methods and Programs in Biomedicine Update*, vol. 1, p. 100036, 2021.

[64] M. A. Khan, T. Akram, Y.-D. Zhang, M. Alhaisoni, A. Al Hejaili, K. A. Shaban, U. Tariq, and M. H. Zayyan, "Skinnet-endo: Multiclass skin lesion recognition using deep neural network and entropy-normal distribution optimization algorithm with elm," *International Journal of Imaging Systems and Technology*, vol. 33, no. 4, pp. 1275–1292, 2023.

[65] M. Rasel, U. H. Obaidellah, and S. A. Kareem, "convolutional neural network-based skin lesion classification with variable nonlinear activation functions," *IEEE Access*, vol. 10, pp. 83 398–83 414, 2022.

[66] A. C. Foahom Gouabou, R. Iguernaissi, J.-L. Damoiseaux, A. Moudafi, and D. Merad, "End-to-end decoupled training: A robust deep learning method for long-tailed classification of dermoscopic images for skin lesion classification," *Electronics*, vol. 11, no. 20, p. 3275, 2022.

[67] S. Shen, M. Xu, F. Zhang, P. Shao, H. Liu, Y. Xu, and L. Ma, "A low-cost high-performance data augmentation approach for skin lesion classification using convolutional neural networks," *Journal of Healthcare Engineering*, 2022.

[68] R. Kaur, H. GholamHosseini, R. Sinha, and M. Lindén, "Melanoma classification using a novel deep convolutional neural network with dermoscopic images," *Sensors*, vol. 22, no. 3, p. 1134, 2022.

[69] V. Dillshad, M. A. Khan, M. Nazir, O. Saidani, N. Alturki, and S. Kadry, "D2lfs2net: Multi-class skin lesion diagnosis using deep learning and variance-controlled marine predator optimisation: An application for precision medicine," *CAAI Transactions on Intelligence Technology*, 2023.

[70] E. S. Nugroho, I. Ardiyanto, and H. A. Nugroho, "Boosting the performance of pretrained cnn architecture on dermoscopic pigmented skin lesion classification," *Skin Research and Technology*, vol. 29, no. 11, p. e13505, 2023.

[71] A. Mehmood, Y. Gulzar, Q. M. Ilyas, A. Jabbari, M. Ahmad, and S. Iqbal, "Sbxception: a shallower and broader xception architecture for efficient classification of skin lesions," *Cancers*, vol. 15, no. 14, p. 3604, 2023.

[72] S. B. Mukadam and H. Y. Patil, "Skin cancer classification framework using enhanced super resolution generative adversarial network and custom convolutional neural network," *Applied Sciences*, vol. 13, no. 2, 2023.

[73] A. Dahou, A. O. Aseeri, A. Mabrouk, R. A. Ibrahim, M. A. Al-Betar, and M. A. Elaziz, "Optimal skin cancer detection model using transfer learning and dynamic-opposite hunger games search," *Diagnostics*, vol. 13, no. 9, 2023.

[74] C. Supriyanto, A. Salam, J. Zeniarja, and A. Wijaya, "Two-stage input-space image augmentation and interpretable technique for accurate and explainable skin cancer diagnosis," *Computation*, vol. 11, no. 12, p. 246, 2023.

[75] T. Khater, S. Ansari, S. Mahmoud, A. Hussain, and H. Tawfik, "Skin cancer classification using explainable artificial intelligence on pre-extracted image features," *Intelligent Systems with Applications*, vol. 20, p. 200275, 2023.

[76] I. A. Ahmed, E. M. Senan, H. S. A. Shatnawi, Z. M. Alkhraisha, and M. M. A. Al-Azzam, "Multi-models of analyzing dermoscopy images for early detection of multi-class skin lesions based on fused features," *Processes*, vol. 11, no. 3, p. 910, 2023.

[77] S. İlkin, T. H. Gençtürk, F. K. Gülağız, H. Özcan, M. A. Altuncu, and S. Şahin, "hybsvm: Bacterial colony optimization algorithm based svm for malignant melanoma detection," *Engineering Science and Technology, an International Journal*, vol. 24, no. 5, pp. 1059–1071, 2021.

[78] M. S. Pitchiah and T. Rajamanickam, "Efficient feature based melanoma skin image classification using machine learning approaches." *Traitement du Signal*, vol. 39, no. 5, 2022.

[79] T. M. Alam, K. Shaukat, W. A. Khan, I. A. Hameed, M. U. Sarwar, and I. A. Hameed, "An efficient deep learning-based skin cancer classification model using ham10000 dataset," *Diagnostics*, 2022.

[80] M. Arshad, M. A. Khan, U. Tariq, A. Armghan, and S. Almotiri, "A computer-aided diagnosis system using deep learning for melanoma detection," *Journal of Healthcare Engineering*, 2021.

[81] I. Iqbal, M. Younus, K. Walayat, M. U. Kakar, and H. Mirza, "Automated multi-class classification of skin lesions using deep learning," *Computers in Medicine and Imaging*, 2021.

[82] I. Abunadi and E. M. Senan, "Deep learning and machine learning techniques of diagnosis dermoscopy images for early detection of skin diseases," *Electronics*, vol. 10, no. 24, p. 3158, 2021.

[83] S. Gopikha and M. Balamurugan, "Regularised layerwise weight norm based skin lesion classification using deep learning," *Computer Systems Science and Engineering*, 2023.

[84] Z. Lan, S. Cai, X. He, and X. Wen, "Fixcaps: An improved capsules network for diagnosis of skin cancer using the ham10000 dataset," *IEEE Access*, 2022.

[85] S. Moldovanu, F. A. D. Michis, K. C. Biswas, and C. Culea, "Skin lesion classification based on surface fractal features," *Cancers*, 2021.

[86] X. Deng, Q. Yin, and P. Guo, "Efficient structural pseudoinverse learning-based method for skin lesion classification," *Complex & Intelligent Systems*, 2022.

[87] D. N. Anggraini Ningrum, S.-P. Yuan, W.-M. Kung, C. H. Tang, and C.-H. Yu, "Deep learning classifier with patient's metadata for improving skin lesion diagnosis," *Journal of Multidisciplinary Healthcare*, 2021.

An Investigation into the Risk Factors of Forest Fires and the Efficacy of Machine Learning Techniques for Early Detection

Asma Cherif[1]*, Sara Chaudhry[2], Sabina Akhtar[3]
Department of Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia[1]
Center of Excellent in Smart Environment Research, King Abdulaziz University, Jeddah, Saudi Arabia[1]
Department of Computer Science, Bahria University, Shangrilla Road, Islamabad, Pakistan[2,3]

***Abstract*—Forest fires are a major environmental hazard that can have significant impacts on human lives. Early detection and swift action are crucial for controlling such situations and minimizing damage. However, the automatic tools based on local sensors in meteorological stations are often insufficient for detecting fires immediately. Machine learning offers a promising solution to forecast forest fires and reduce their rapid spread. In recent state-of-the-art solutions, only one or two techniques have been utilized for prediction. In this research, we investigate several methods for forest fire area prediction, including Long Short Term Memory (LSTM), Auto Regressive Integrated Moving Average (ARIMA), and Support Vector Regression (SVR). Our aim is to identify the most effective and optimal method for predicting forest fires. After comparing our results with other artificial intelligence and machine learning techniques applied to the same dataset, we found that the LSTM approach outperforms the ARIMA and SVR predictors by more than 92%. Our findings also indicate that the LSTM algorithm has a lower estimation error when compared to other predictors, thus providing more accurate forecasts.**

***Keywords*—Machine Learning; Forest Fire; LSTM; ARIMA; SVR**

I. Introduction

Forests play a crucial role in the earth's ecosystem and environmental sustainability. Wildland fires, commonly known as forest fires, are among the deadliest disasters that pose a threat to forest preservation, causing ecological devastation and casualties [1], [2]. In recent decades, human activities have caused the number of forest and land fires to significantly increase [3]. The 2019 wildfire in southeast Australia is an example of the devastating impact of forest fires, destroying over 11.2 million hectares of forest and leading to the extinction of numerous creatures [4], [5]. Detecting and controlling forest fires is becoming increasingly challenging. Rapid detection is essential to effective control. Detection techniques typically include smoke detection, satellite monitoring, and local perception (such as data analysis). While satellite monitoring is costly and subject to delays, smoke detection requires expensive equipment and maintenance. Data analysis, in contrast, is a less expensive and more accessible method for detecting and analysing forest fires [6], [7].

Human carelessness and natural factors like lightning are the two primary causes of forest fires. Recent studies have indicated that forest fires and climate change are related [8]. Inconsistencies in weather patterns, such as irregular rainfall, wind patterns, temperature swings, and precipitation, have contributed to an increase in forest fires in recent years [9]. The amount of money spent by the government to fight forest fires has also increased. Ecologists have employed several methods to understand how the forest landscape is changing, and statistical models have proven to be useful in examining how the patterns of forest fires are evolving for a particular location. However, current technologies cannot accurately predict the site of forest fires based on past data and environmental circumstances. Detecting forest fires early can help contain and reduce the scope of a disaster, and such solutions can aid firefighters in their efforts.

Artificial intelligence has emerged as a promising tool for predicting wildland fires with high accuracy [10], [11]. Neural networks (NN) are commonly used to forecast the occurrence of wildland fires and can reduce false alarms with the aid of infrared data [12]. Support vector machines (SVM) have also proven to be effective in increasing the accuracy and effectiveness of wildland fire predictions [13], [14]. Random Forest (RF) [15], [16] is considered to be a technique that performs quite well in these scenarios and its hybrid with SVM that is Random vector forest regression (RVFR) has also been studied to bring better results [17]. Additionally, data mining techniques such as logistic regression [18], [19], decision tree (DT) [15], and fuzzy logic [16] have been used to develop wildland fire prediction models. These techniques can help authorities detect forest fires early, allowing for more efficient and effective containment efforts.

Forest fires can cause significant losses to nature, the environment, and property. To prevent these losses, it is essential to have accurate and timely forest fire prediction systems in place. However, relying solely on automatic tools based on local sensors in meteorological stations may not be enough to instantly detect fires. Therefore, studying and selecting a suitable model for forest fire prediction can play a vital role in preventing fires from occurring or spreading. By accurately predicting forest fires, authorities can take proactive measures to ensure the safety of people, property, and the environment.

The main objective of this research is to identify the factors that contribute to the rapid spread of forest fires, including wind speed, humidity, temperature, precipitation, FFMC (Fine Fuel Moisture Code), DMC (Duff Moisture Code), DC

*Corresponding authors.

(Drought Code), and ISI (Initial Spread Index). The aim is to develop a predictive model that can assist authorities in assessing the long-term impacts of climate change. Furthermore, the study compares the outcomes with other machine learning and artificial intelligence techniques that were previously applied to the same dataset and reported in the literature. However, it is worth noting that only a limited number of techniques were utilized in the state-of-the-art solutions. Therefore, this research aims to fill this gap by applying various machine learning algorithms and selecting the best prediction model. By doing so, it can contribute to the development of a more accurate and effective forest fire prediction model that can help prevent losses due to forest fires.

This study is focused on addressing the following research questions:

- RQ1: What are the current state-of-the-art techniques and models used for predicting wild fires, as reported in the literature?
- RQ2: Which specific features or variables are most influential in accurately predicting the spread of forest fires?
- RQ3: Based on a thorough review of previous literature, what are the most suitable models that can be trained for our specific dataset?
- RQ4: What methods can be employed to improve the accuracy of the prediction model, such as reducing the root mean square error or mean absolute error?

The rest of the paper is structured in the following way: Section II provides a comprehensive literature review. Sections III and IV elaborate on the methodology and results, respectively. Finally, the paper concludes with a summary of findings and future research directions in the last section.

II. Literature Review

Numerous studies and research have been conducted to predict forest fires and minimize their damage.

Cortez et al. [14] proposed data mining algorithms, including support vector machine (SVM) and random forest, which utilized four different features (spatial, temporal, weather attributes, and FWI* component) to predict the burned area of a forest. However, this proposed solution based on SVM can only predict small fires and has a low accuracy rate for large fires.

Elshewy et al. [20] proposed three machine learning algorithms, which were applied to a dataset of 517 entries and 13 features per entry. The algorithms were tested in two scenarios: one with the entire dataset used for training and testing, and the other with 70% of the attributes used for training and 30% for testing. It was concluded that linear regression had the highest accuracy score of 0.99, surpassing the other two algorithms.

In their study, Nebot et al. [21] utilized fuzzy logic models to predict the burned area of forest fires. The authors employed two powerful fuzzy systems, the Adaptive Neuro-Fuzzy Inference System (ANFIS) and Fuzzy Inductive Reasoning (FIR), to model burned sections of fires in Portugal. To validate their models, a 10-fold cross-validation method was used, involving model identification and validation repeated 10 times. The authors created 100 FIR and 100 ANFIS models to test the generalization performance of these hybrid fuzzy models. The FIR models exhibited the highest predictive power and the lowest MAE and RMSE errors when compared to all other models.

In their study, Al Janabi et al. [22] explored the use of soft computing algorithms for predicting forest fires. They collected over 500 entries for Montesinho Natural Park (MNP) in Portugal, which included 12 spatial and temporal parameters, a fire weather indicator, and burned area. To extract significant insights from the data, the researchers utilized Principle Component Analysis and Particle Swarm Optimization techniques. They employed five strategies simultaneously to compare and select the optimal solution. The SVM strategy with minimal estimating error was found to be the most successful approach.

Liang et al. [23] investigated the use of meteorological variables, including temperature, humidity, and precipitation, in combination with various predictive models for a fire-prone region in Alberta, Canada. The data was obtained from the National Fire Database of Canada. To estimate the size of forest fires, the researchers employed Back Propagation Neural Networks (BPNN), Recurrent Neural Networks (RNN), and Long Short Term Memory Models. The LSTM model exhibited the highest performance with an accuracy rate of 90.9%.

Zhang et al. [24] developed a Convolutional Neural Network (CNN) with a complex architecture for predicting forest fire susceptibility in Yunnan Province, China. The researchers manually assembled pictures to study the effects of various parameters. To optimize the model, they employed multicollinearity analysis and the Information Gain Ratio (IGR) approach. The model architecture was inspired by Google's AlexNet [25] and categorized photos into different classes. The CNN model exhibited a strong predictive potential with an AUC of 0.86.

Long Short-Term Memory (LSTM) algorithm has been extensively used to address time series prediction problems due to its high accuracy and speed [26]. The LSTM model is particularly advantageous in predicting trends, making it well-suited for predicting wildland fire burned areas. However, the complexity of factors and computations involved has limited the potential of the LSTM model in this domain. Therefore, further exploration and development of an enhanced LSTM model is necessary to improve the accuracy and applicability of predicting burned regions in wildland fires.

Li and Huang [27] proposed a modified model based on the Long Short-Term Memory (LSTM) network with multiple input layers and an attention mechanism module to predict the burned regions in wildland fires. The study used the Montesinho dataset and drew some interesting conclusions. First, to reduce computational complexity and interference, correlation analysis was used to identify potential related elements, and unnecessary factors were removed. This helped determine the various causes of wildfire speed. Second, a Multi-AM-LSTM

*Fire Weather Index (FWI) system was introduced in 1970s. It only required readings of four meteorological observations (i.e. temperature, relative humidity, rain and wind).

model was developed to learn effective features and predict burned areas. The proposed model achieved an impressive accuracy rate of over 96%. Comparison with other models showed that LSTM with an attention mechanism is particularly effective in this context.

George E. [11] proposed a novel mechanism for predicting forest fire risk using only meteorological data, independent of weather forecasting systems. The study utilized support vector machines to achieve a high accuracy rate of up to 96% for August in a two-class prediction of fire risk. The findings demonstrate the potential of this approach to accurately predict forest fire risk, which can aid in preventing and mitigating the damage caused by wildfires.

In their study, Richa.S et al [28] employed a variety of machine learning techniques to test eight classification models using the Cortez Morais dataset, which includes 517 instances and 13 attributes. The algorithms were evaluated using several metrics, including precision, recall, f-score, accuracy, and Area Under the Curve (AUC). The Boosted Decision Tree model achieved an accuracy of 72%, which outperformed the neural network model by 6%, the 2 Class Bayes machine by 14%, and the remaining five algorithms by 3%. These results suggest that the Boosted Decision Tree model may be a promising approach for predicting forest fires based on the Cortez Morais dataset.

Ahmed Al Janabi's study [20] utilized particle swarm optimization (PSO) to segment fire zones and principal component analysis (PCA) to identify key patterns or clusters. The study then employed five soft computing (SC) techniques based on neural networks in parallel to determine the best method for predicting forest fires with accuracy and optimality. The performance of these predictors was evaluated using five quality metrics, including information gain, relative absolute error, mean absolute error, and root mean squared error (RMSE). The results indicated that the SVM technique outperformed RBF, MPNN, PNN, and CCN predictors in terms of both effectiveness and efficiency. These findings suggest that the SVM technique may be a promising approach for predicting forest fires based on PSO and PCA segmentation.

In recent years, several projects have investigated the application of various fire detection methodologies and technologies. For instance, a study by ZQ et al. [6] presented an overview of the different forest fire detection methods that have been introduced, including wireless sensor networks, optical sensors, digital cameras, and satellite imaging. Similarly, Gibson et al. [5] explored the use of statistical methods in predicting forest fires while considering the various factors and challenges associated with it. Hong et al. [29] examined the use of different sensors, such as temperature, smoke, flames, and flammable materials, in fire detection systems.

David et al. [25] proposed an efficient machine learning method for predicting total burned areas for specific wildfire episodes with high accuracy. The transparent open box (TOB) algorithm does not rely on regression, correlation, or statistical distribution assumptions, nor does it use any hidden layers or complex calculations. This method can provide valuable insights to aid in mitigating specific burn events as they occur, resulting in various short- and long-term benefits. Furthermore, the proposed strategy can be adapted to anticipate and analyze data from other agricultural systems that rely on complex relationships between meteorological and environmental factors, in addition to regression and correlation-based approaches.

III. Methodology

This section provides a comprehensive overview of the dataset utilized in the study, along with the preprocessing steps taken to predict fire and details of the models used. First, we will discuss the dataset in detail, followed by a description of the preprocessing techniques employed to ensure the accuracy and reliability of the predictions. Finally, we will delve into the specific models utilized in this research, highlighting their strengths and limitations in predicting fire occurrences.

The workflow for our study is presented in Fig. 1. The diagram outlines the various steps taken to analyze the dataset, preprocess the data, and develop and evaluate the models used in the study. Each step in the workflow is described in detail in the subsequent sections.

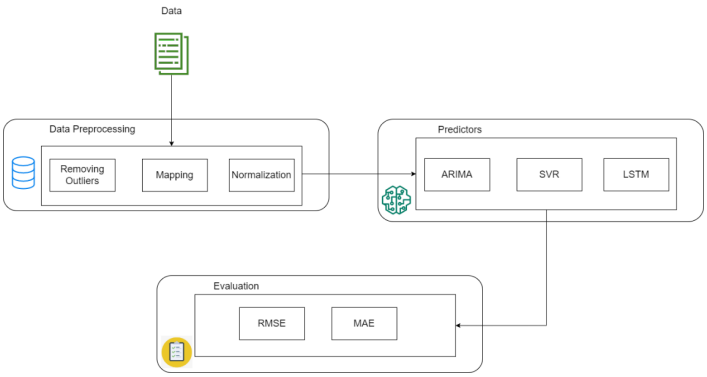


Fig. 1. Workflow process.

A. Dataset

The dataset used in this research was sourced from the UCI machine learning repository, consisting of 517 unique entries recorded at various periods between January 2003 and December 2020. The dataset comprises 12 attributes categorized into three groups: weather conditions, fire weather index (FWI), and geographical and temporal components, with the overall burned area serving as the output feature. The data was collected from two sources, as described in [7]. The first source recorded details such as date, time, location within a 9-by-9 grid, vegetation type, FWI elements, and burned area each time a forest fire occurred. The FWI, a Canadian system for categorizing fire hazard that was used to indicate the intensity of the fire. The second source consisted of weather measurements taken every 30 minutes by a meteorological station. The two databases were combined to create a single dataset. Table I provides a detailed explanation of each attribute.

B. Data Exploration

The dataset contains 517 rows of data. To gain insights into the distribution of the numerical variables, we created boxplots for each variable. Upon analysis, we identified the presence of outliers in some of the variables. The boxplots of these variables are presented in Fig. 2.

TABLE I. DESCRIPTION OF DATASET

Attributes	Description
Spatial Attribute	
X	Coordinates on x-axsis (1-9)
Y	Coordinates on y-axsis (1-9)
Temporal Attributes	
Month	January to December months start
Day	Day starts with Monday and ends on Sunday
Intensity of Weather Attributes	
FFMC	FFMS code in digits
DMC	DMC code in digits
DC	DC code in digits
ISI	ISI code in digits
Weather Attributes	
Temperature	Temperature in degree Celsius
RH	RH in percentage
Rain	Rain in mm
Wind	Relative humidity in percentage
Target	
Area	Total area burnt in hectare

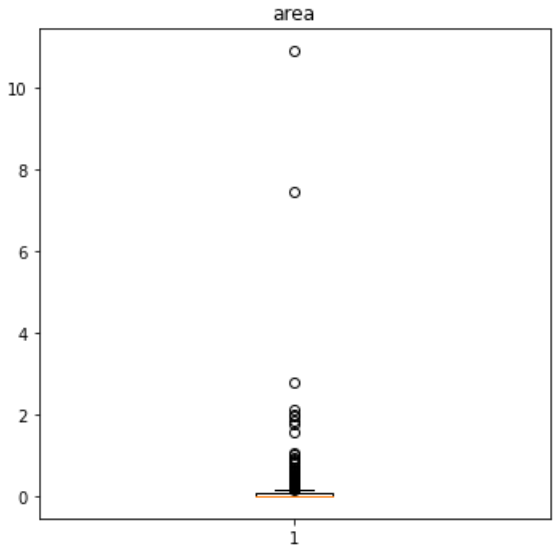


Fig. 2. BoxPlot of area.

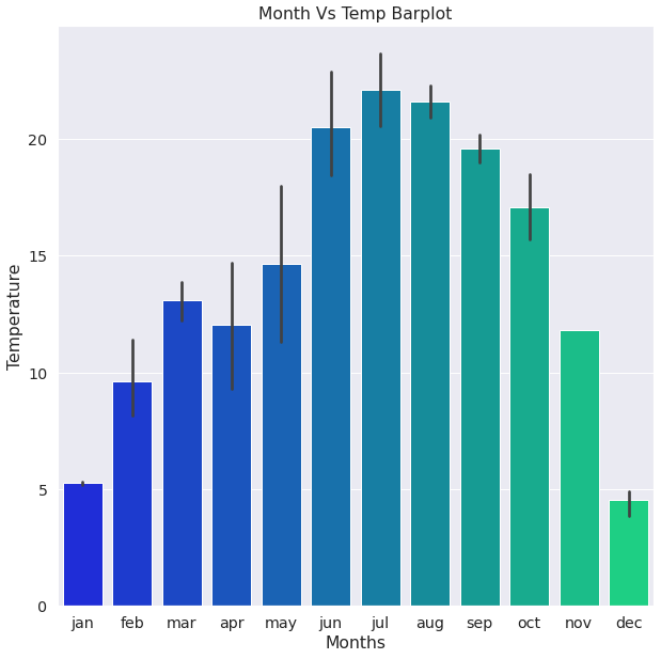


Fig. 3. Month vs Temperature.

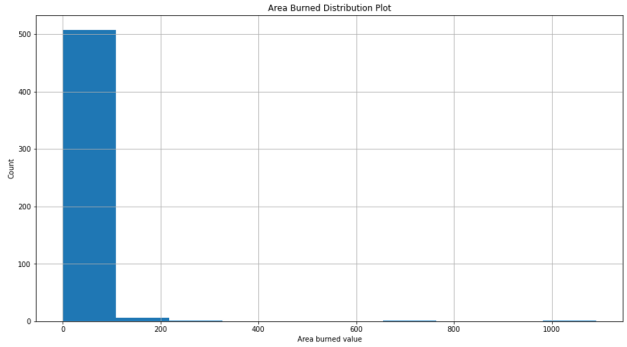


Fig. 4. Area distribution.

Our target variable is the "Area" variable which also has a significant number of outliers. However, we have decided to remove only the top 10% of outliers from the right side of the data. After removing these outliers, we were left with a total of 515 rows, down from the initial 517 rows.

The graph in Fig. 3 highlights that the months of June, July, August and September experience the highest temperatures. This phenomenon is primarily due to the summer season, resulting in high temperatures and increased humidity levels. As September concludes, the temperature starts to decline gradually.

1) Target transformation: The variable "area" has a highly skewed distribution that is positively skewed towards 0, as shown in Fig. 4. Modeling such a skewed distribution during training impacts the model performance. Ideally, the variable should have a normal distribution. To transform "area" into a normal distribution range, we performed a log transformation as shown in Fig. 5.

2) Mapping: To apply different models to our dataset, we needed to convert our categorical variables into numeric variables. Specifically, we mapped the "month" attribute to its corresponding month number and the "day" attribute to its corresponding day number, with Sunday being the first day. Additionally, we utilized one-hot encoding to apply LSTM to our dataset.

3) Normalization: The feature values in the dataset exhibited a wide range, potentially leading to increased computing complexity and inaccurate predictions. Thus, normalization is necessary to assign equal weight to each variable. This involves utilizing the Min-Max Scaling technique to quantify the variables, linearly transforming the original variable ranges to new ranges.

$$Z = \frac{x_i - min(x)_i}{max(x)_i - min(x)_i} \quad (1)$$

In Eq. 1, Z is the output value, x_i is the value of the

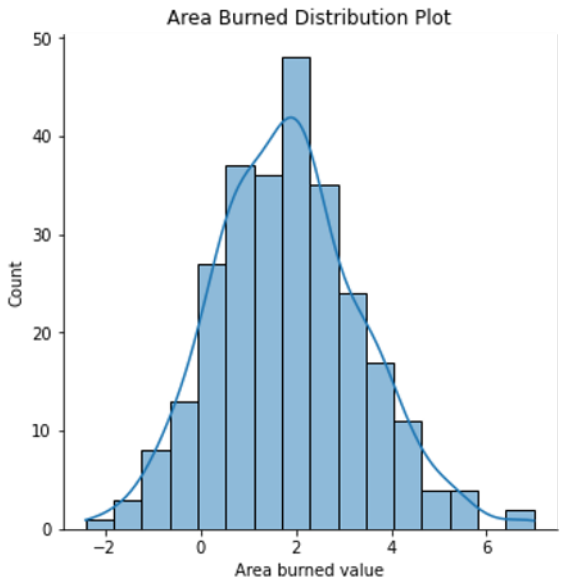


Fig. 5. Transformed area distribution.

variables, max and min denote the minimum and maximum values of each variable, respectively.

C. Feature Selection

Feature selection is a technique employed to improve accuracy in the machine learning process. Additionally, it enhances the predictive ability of algorithms by choosing the most essential variables while excluding redundant and irrelevant ones. In this research we applied two techniques: the Mutual Information (MI) and correlation analysis.

D. Mutual Information

MI is a measure of the amount of information that one random variable can provide about another random variable. It is often used for dimension reduction [30].

A formal way to express the mutual information between two random variables X and Y is as follows:

$$I(X,Y) = H(X) - H(X \mid Y) \tag{2}$$

Where $I(X,Y)$ is the mutual information for X and Y, $H(X)$ is the entropy for X and $H(X \mid Y)$ is the conditional entropy for X given Y.

Fig. 6 displays the results of applying the MI to our features and target. It indicates that the month has a great mutual information with the target feature "area". Besides, the temp attribute is showing some mutual information with the target. Thus, the events of fire have a relation with temperature.

E. Correlation Analysis

Correlation analysis is used to check which are the most important variables in the data set. The correlation between these characteristics is applied using the following equation:

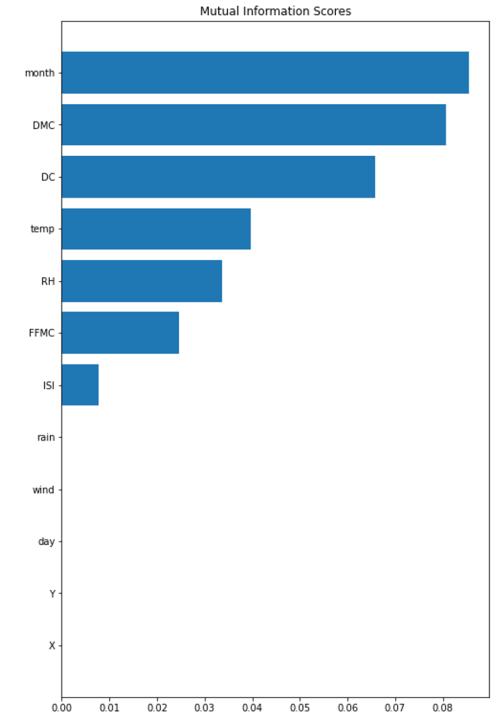


Fig. 6. The Mutual inclusion between the features and the target.

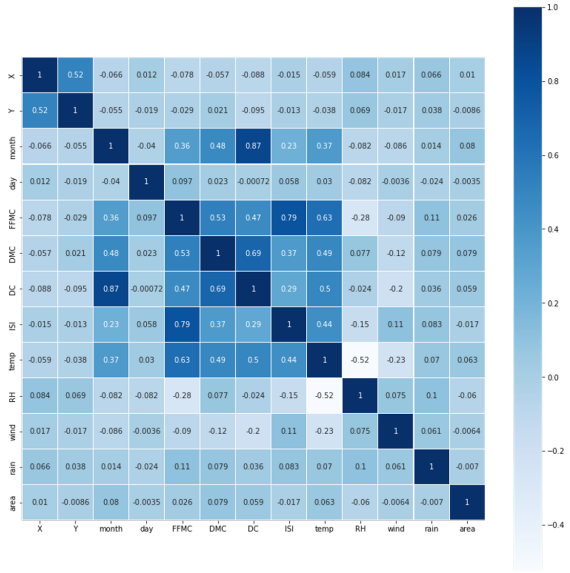


Fig. 7. Correlation analysis.

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 (\sum x)^2][n\sum y^2 (\sum y)^2]}} \tag{3}$$

where n represents total observation and $-1 < r < 1$, the positive sign $(+)$ means positive correlation and negative sign $(-)$ means negative relationship between variables.

Fig. 7 indicates that 5 features have some correlation with the target which are: Month, DC, DMC, RH, Temp, ISI.

F. Predictors

In the following discussion, we will examine the predictors utilized for forecasting the area.

1) Long short term memory: Particular types of recurrent neural networks can be trained to identify long-term dependencies in data. As such, we utilized an LSTM network to address prediction problems using a dataset sourced from UCI Machine Learning respiratory. The LSTM network incorporates a cell state through which information is passed, and the addition or removal of information is regulated by input gates, forget gates, and output gates. To gain insight into the functioning of the LSTM network and how the gates acquire pertinent information, refer to Fig. 8[†].

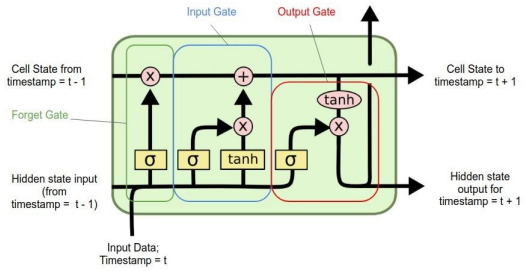


Fig. 8. LSTM network.

a) Forget gate: The forget gate plays a crucial role in determining which information should be kept or removed from the cell state. This gate processes both current input and relevant information using a sigmoid function, resulting in values that fall between 0 and 1. A value of 1 indicates that the information is to be retained, while a value of 0 indicates that it should be forgotten.

b) Input gate: The input gate is responsible for deciding which data to preserve in the cell state. This gate processes both input and hidden state information through a hyperbolic tangent (tanh) function. The resulting output is multiplied by a sigmoid function to determine whether to keep or discard the information. If the sigmoid output approaches 1, the information is retained, whereas a sigmoid output approaching 0 results in the information being discarded.

c) Output gate: The output gate plays a critical role in determining the final output of the cell state. This gate filters the cell state output to produce a refined version of the information that should be included in the final output.

2) ARIMA: ARIMA, short for AutoRegressive Integrated Moving Average, is a widely-used model in time series data analysis. It utilizes historical values of a time series to forecast future values. However, external factors can also have an impact on the time series and serve as accurate predictors of future values. This is where the SARIMAX model comes into play, by introducing an exogenous variable X. In statistics, predictors or input variables are referred to as exogenous, while the target variable is referred to as endogenous. The SARIMAX model enhances the $SARIMA(p,d,q)(P,D,Q)_m$ model by incorporating the effects of exogenous variables, producing more accurate forecasts [31]. The parameters p,d,q are trend elements which refer to autoregressive order, difference order and moving average order respectively. Similarly, the parameters P,D,Q are seasonal elements refering to autoregressive order, difference order and moving average order. The parameter m indicates seasonal length.

In Eq. 4, we may easily add any number of exogenous variables X_t to the $SARIMA(p,d,q)(P,D,Q)_m$ model to reflect the present value Y_t.

$$Y_t = SARIMA(p,d,q)(P,D,Q)_m + \sum_{i=1}^{n} \beta_i x_t^i \qquad (4)$$

Eq. 4 indicates that the SARIMA model is a linear combination of the historical values of a time series and the corresponding error terms. The SARIMAX model, an extension of SARIMA, incorporates exogenous variables, transforming it into a more comprehensive linear model. This extension allows to model the relationship between the time series and external factors in a more accurate and effective manner, thereby yielding improved forecasting results.

3) Support Vector Regression SVR: Support Vector Regression (SVR) is an extension of the Support Vector Machine (SVM) algorithm, designed to handle regression problems in both linear and nonlinear domains. While SVM is primarily used for classification tasks with categorical variables, SVR is tailored for continuous variable prediction. This fundamental distinction enables SVR to model complex relationships between variables and generate accurate predictions from continuous datasets.

Despite a few minor differences, SVR operates on the same fundamental principles as SVM. Both algorithms aim to locate a curve that fits the given data points. However, as a regression procedure, SVR utilizes this curve to match the vector to the point of the curve, rather than as a decision boundary. This matching process is facilitated by the use of support vectors, which helps to identify the closest match between the data points and the function used to represent them.

IV. Results

This section presents the results of three models that were developed to predict forest fire areas based on historical data. By analyzing the performance of each model, we can gain insights into the effectiveness of their respective methodologies and techniques. Through these evaluations, we aim to identify the most accurate and reliable model that can be used to predict forest fire areas with the highest level of confidence.

A. LSTM

LSTM has been widely used to address time series prediction problems due to its high accuracy and speed. Here, we explain the development of the prediction model using LSTM.

[†]Figure from: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

1) Data Splitting: After preprocessing our data for use in LSTM, we were left with 514 rows and 30 features out of a total of 517 rows. To predict the fire area for the next day, we implemented a window length of 10. During the training phase, we utilized 80% of data, incorporating all features, and reserved 20% rows for testing. This model will be used to forecast the fire area for the next 38 days.

2) Model Training: Several models have been trained with different parameters as shown in Table II.

TABLE II. TRAINED LSTM MODELS

Model No.	Layers	1^{st} Layer Neurons	2^{nd} Layer Neurons	3^{rd}Layer Neurons
1	1	100		
2	1	50		
3	2	50	55	
4	3	100	100	55

All models utilize the ReLU activation function. Additionally, we have included a dense layer with eight neurons, as illustrated in Fig. 9. To optimize our model, we employed the Adam optimizer with default settings for the learning rate (lr=0.001) and weight decay (0.9). During training, we used a batch size of 25 and conducted 50 epochs to ensure that our models are able to learn from the data effectively.

Fig. 9 illustrates the second model summary, which consists of a total of 78,368 parameters, all of which have been trained.

```
Model: "Forest_fire_Prediction_Model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 LSTM_Hidden_Layer_1 (LSTM)  (None, 10, 100)           43600

 LSTM_Hidden_Layer_2 (LSTM)  (None, 55)                34320

 output_layer (Dense)        (None, 8)                 448

=================================================================
Total params: 78,368
Trainable params: 78,368
Non-trainable params: 0
_________________________________________________________________
```

Fig. 9. Model 2 summary.

To fine-tune the hyperparameters, we utilized the validation set. The hyperparameters for the proposed bidirectional LSTM models are presented in Table III.

TABLE III. HYPERPARAMETER OF LSTM

Hyperparameter	Values
Activation Function	ReLU
Number of LSTM layer	1,2,3,4
Dropout	Yes
Regularization	Yes
Optimizer	Adam
Learning rate	0.001
Batch size	25
Epochs	50

Loss indicates the performance of the model whether it is able to predict accurately or not. If loss is zero then it is accurate. Fig. 10 displays the loss with respect to the number of epochs for all the trained models. As shown in the figure, the loss decreases as the number of epochs increases. Besides, Models 3 and 4 demonstrate a better convergence between the training and validation losses.

Fig. 11 explains the ground truth or the actual area versus the predicted area for the models. We have predicted occurrence of forest fire area for thirty eight (38) days.

B. ARIMA

ARIMA is an extensively used model for analysis of time series data. It utilizes historical values of a time series to forecast future values. In this section, we explain how we prepared the data for creating a model using ARIMA and explain our findings.

1) Data Splitting: After preprocessing, we were left with 515 out of 517 rows and 13 features for ARIMA analysis. We trained the model on 477 days of data using all features, and reserved 38 rows for testing. The ARIMA model forecasted the area of fire for the next 38 days, as illustrated in Fig. 12.

2) SARIMAX summary: The SARIMAX model is utilized when an exogenous variable 'X' is included. In statistics, the term *exogenous* is used to describe predictors or input variables, while *endogenous* refers to the target variable- what we aim to predict. For ARIMA training, we determine the values of (P, D, Q). To determine these values for our dataset, we used the auto arima built-in function from the pdarima library, resulting in P=0, Q=1, and D=4. We also attempted to predict forest fire occurrences for the next 38 days, and Fig. 13 displays the SARIMAX model's results.

C. SVR

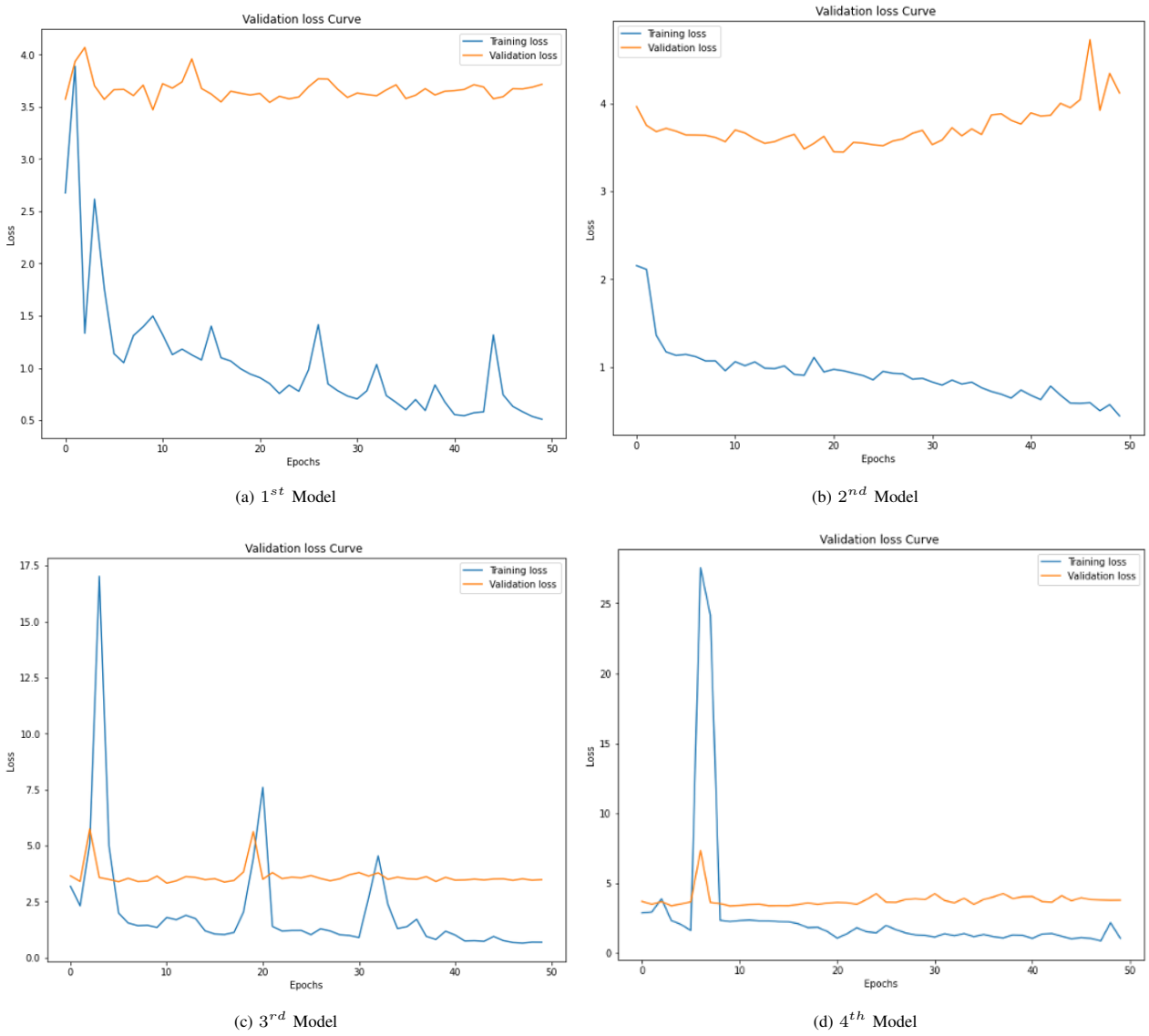
In SVR, we began with a total of 517 rows. After preprocessing, we were left with 515 rows and 13 features. For training, we used 477 days of data, including all features. The remaining 38 rows were used for testing. The model is designed to forecast the area of fire for the next 38 days. We utilized a linear kernel with C=1 and gamma of 2e-5 for the SVR. The SVR model forecasted the area of fire for the next 38 days, as illustrated in Fig. 14.

D. Overall Comparison of Models

To assess the model's effectiveness, we employ the RMSE (root mean square error) metric defined in Eq. 5. These error measurements are widely used in the field of data analysis and are valuable tools for determining the accuracy of a model's predictions. By analyzing the RMSE, we can gain insight into the model's strengths and weaknesses and identify areas for improvement.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|^2} \tag{5}$$

Based on the data presented in Table IV, it is evident that the ARIMA and SVR models have the highest RMSE error. Therefore, we can conclude with confidence that LSTM can be utilized for predicting fires in this dataset or other similar datasets. Indeed, it outperforms the two other method by more than 92%.



(a) 1^{st} Model

(b) 2^{nd} Model

(c) 3^{rd} Model

(d) 4^{th} Model

Fig. 10. Training loss for 4 LSTM models.

TABLE IV. COMPARING THE PERFORMANCE OF THE TESTED MODELS

Model	RMSE
ARIMA	16.59
LSTM 1	1.219
LSTM 2	1.233
LSTM 3	**1.15**
LSTM 4	1.16
SVR	18.82

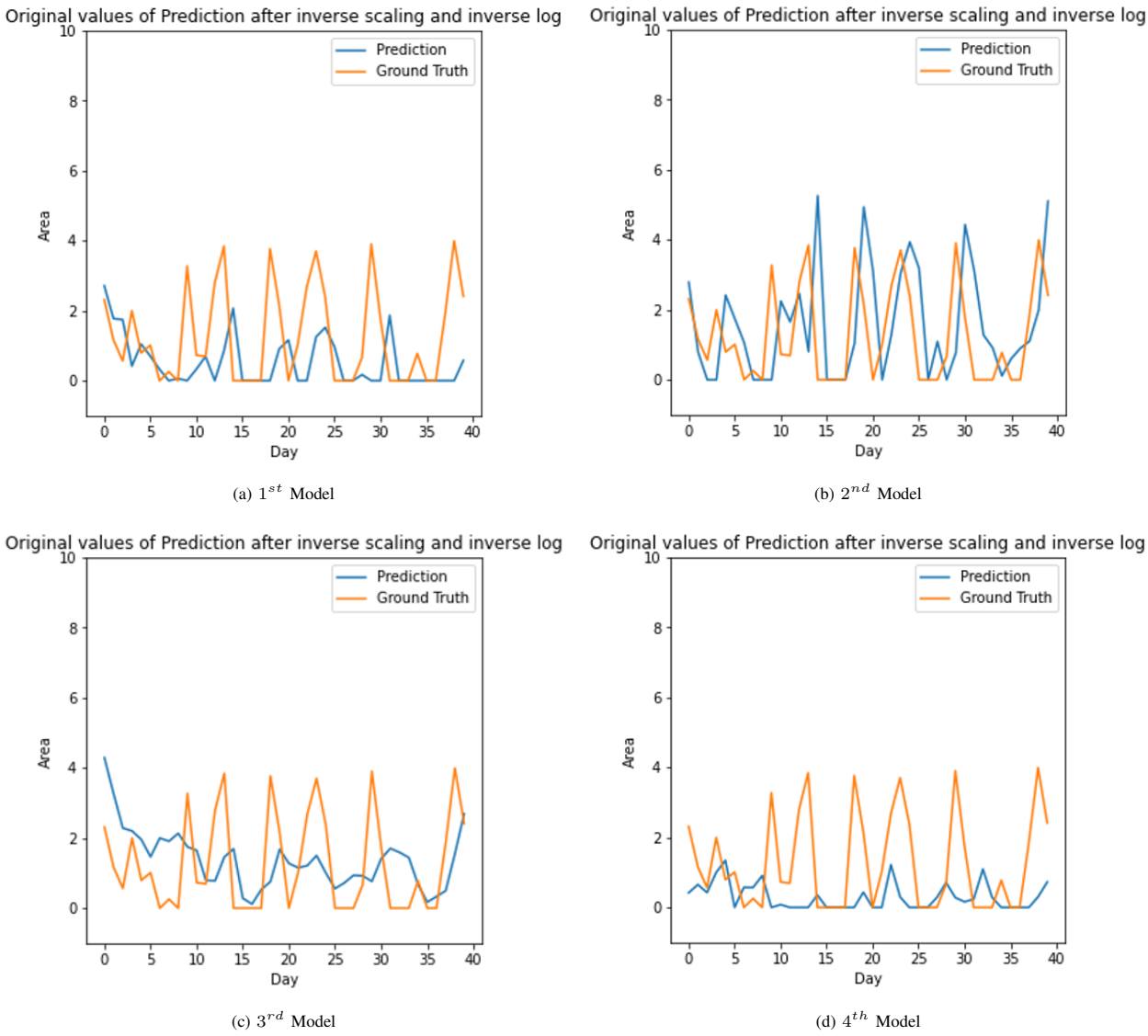
E. Discussion

To determine the predictive accuracy of artificial intelligence and machine-learning models, it is essential to compare their outcomes with those of other similar models. This is why this section showcases the results of previously published research that was conducted on the same datasets. For this comparison, we use the RMSE error, which is depicted in Eq. 5. By comparing these errors with those of other models, we can assess the effectiveness of our own model.

Several models have been applied and evaluated on the same datasets in the literature [7], [22], [21], [32]. Table V summarizes the assessment metric results of these models, along with the errors produced by LSTM, SVR, and ARIMA (shown at the bottom of the table). These datasets have shown promising results for both statistical models and other machine learning techniques. In this study, we have utilized comparable data from previous research to evaluate the performance of our models.

V. CONCLUSION AND FUTURE WORK

The aim of this study was to create an effective machine learning algorithm for predicting the spread of fire in hazardous incidents. This would enable workers to take immediate action, thereby reducing economic and natural losses. The study utilized data on forest fires from Portugal's Montesinho Natural



(a) 1^{st} Model

(b) 2^{nd} Model

(c) 3^{rd} Model

(d) 4^{th} Model

Fig. 11. Prediction of the 4 LSTM models.

Park, which is accessible through the UCI machine-learning repository. After thorough research, the study identified the most suitable machine learning and statistical models, including Long Short Term Memory (LSTM), Auto Regressive Integrated Moving Average (ARIMA), and Support Vector Regression (SVR). The study evaluated the models using the Root Mean Square Error (RMSE), and found that LSTM was the most effective model, producing the best results with an RMSE of 1.15.

The prediction results of LSTM, SVR, and ARIMA were compared to those published in the literature [7], [22], [21], [32]. The methodologies used in these papers included various types and structures of neural networks (NN), Support Vector Machines (SVMs), Decision Trees (DTs), Multiple Regression (MR), Random Forest (RF), Transparent Open-Box (TOB) Network, Radial Basis Function (RBF), Standard Genetic Programming Decision Trees (ST-GP), Linear Regression (LR), Adaptive Neuro-Fuzzy Inference System (ANFIS), Probabilistic Neural Networks (PNN), and Finite Impulse Response (FIR). Using the same data, the comparison aimed to evaluate the performance of the aforementioned methods and to identify the most accurate model for prediction.

The results indicate that the top-performing LSTM models exhibited the lowest RMSE error, as well as the highest predictive accuracy compared to all other models. To enhance accuracy in the future, we can make necessary parameter adjustments and augment data sources by incorporating various factors, such as forest vegetation, cover, types of trees, and Buildup Index. By integrating geographic information system (GIS) data and satellite imagery, we can further optimize this model and achieve even greater precision. This approach can significantly enhance our ability to analyze and understand forest dynamics and provide valuable insights for forest management and conservation efforts.

TABLE V. COMPARISON OF DIFFERENT METHODS

Method	Reference	Parameters	MSE
DT	Cortez et al. [7]	Reduction of the sum of squares	64.5
SVM	Al-Janabi et al. [22]	kernel RBF; C = 84.17; E = 0.001; G = 3800.3	54
RBF	Al-Janabi et al. [22]	Num. Neur. = 100; Min. Rad. = 0.01; Max. Rad. = 519.669; Min. Lambda = 0.01328; Max.Lambda = 9.953	54.2
TOB	David et al. [32]	Two stages; Wn = 0.54; Q = 10; Evol 2-6; Optimum	63.26
ANFIS	Angela Nebot et al. [21]	Hybrid algorithm constant funtion; 50 epochs	64.6
PNN	Al-Janabi et al. [22]	Gaussian kernel; p1 = 31.31; p2 = -27.31; p3 = 1.14; p4 = 6.31	63.2
FIR	Angela Nebot et al. [21]	EWP-EFP; 2-3 FS per variable	48.9
ARIMA	Our Research	P=0, q=1 ,d=4	16.59
SVR	Our Research	Kernal= linear, C=1, Gamma= 2e-5	18.82
LSTM	Our Research	Epoch=50, Batch Size= 25	1.15

SARIMAX Results

Dep. Variable:		y	No. Observations:		477	
Model:	SARIMAX(0, 1, 4)		Log Likelihood		-1.673	
Date:	Sun, 03 Jul 2022		AIC		31.346	
Time:	18:34:04		BIC		89.662	
Sample:	0		HQIC		54.277	
	- 477					
Covariance Type:	opg					
	coef	std err	z	P>|z|	[0.025	0.975]
X	0.0068	0.009	0.743	0.458	-0.011	0.025
Y	-0.0307	0.013	-2.424	0.015	-0.056	-0.006
month	0.0142	0.023	0.609	0.542	-0.031	0.060
day	-0.0031	0.006	-0.482	0.630	-0.016	0.009
DC	-0.0001	0.000	-0.638	0.524	-0.000	0.000
ISI	-0.0047	0.004	-1.074	0.283	-0.013	0.004
temp	0.0064	0.004	1.539	0.124	-0.002	0.015
wind	0.0108	0.011	0.999	0.318	-0.010	0.032
rain	-0.0967	0.143	-0.675	0.500	-0.378	0.184
ma.L1	-0.6290	0.019	-32.575	0.000	-0.667	-0.591
ma.L2	-0.0179	0.031	-0.580	0.562	-0.079	0.043
ma.L3	-0.0725	0.059	-1.230	0.219	-0.188	0.043
ma.L4	-0.0465	0.058	-0.800	0.424	-0.160	0.067
sigma2	0.0622	0.002	40.627	0.000	0.059	0.065

Ljung-Box (L1) (Q):	0.23	Jarque-Bera (JB):	41216.98
Prob(Q):	0.63	Prob(JB):	0.00
Heteroskedasticity (H):	44.06	Skew:	3.65
Prob(H) (two-sided):	0.00	Kurtosis:	48.00

Fig. 13. Sarimax summary.

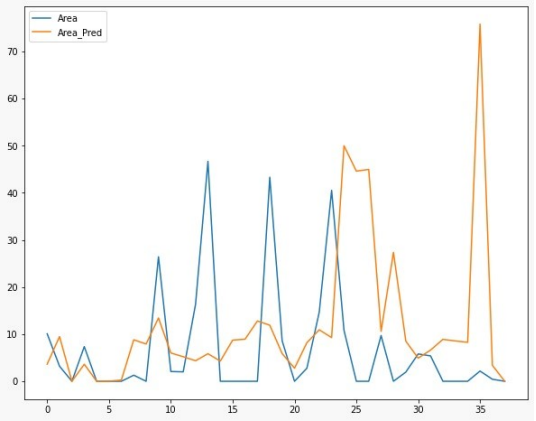


Fig. 14. Actual Area vs Predicted area.

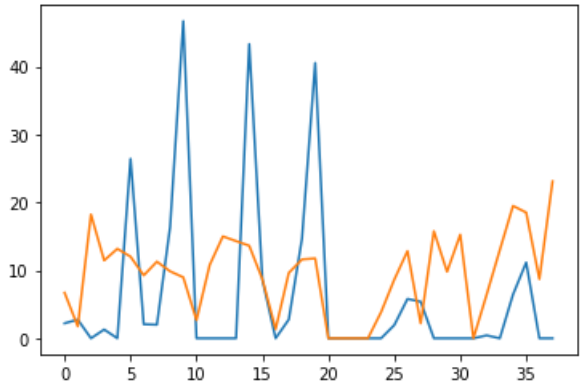


Fig. 12. Area vs Predicted area.

ACKNOWLEDGMENT

This research work was funded by Institutional Fund Projects under grant no. (FPIP:1564-612-1443). The authors gratefully acknowledge technical and financial support provided by the Ministry of Education and King Abdulaziz University, Deanship of Scientific Research, Jeddah, Saudi Arabia.

REFERENCES

[1] D. G. Neary, K. C. Ryan, and L. F. DeBano, "Wildland fire in ecosystems: effects of fire on soils and water," *U.S. Department of Agriculture, Forest Service*, vol. 4, p. 250, 2005.

[2] K. Nyongesa and H. Vacik, "Fire management in mount kenya: A case study of gathiuru forest station," 06 2018.

[3] "Mono-temporal and multi-temporal approaches for burnt area detection using sentinel-2 satellite imagery (a case study of rokan hilir regency, indonesia)," *Ecological Informatics*, vol. 69, p. 101677, 2022.

[4] M. Storey, O. Price, J. Sharples, and R. Bradstock, "Drivers of long-distance spotting during wildfires in south-eastern australia." *International Journal of Wildland Fire*, vol. 29, pp. 459–472, 2020.

[5] R. Gibson, T. Danaher, W. Hehir, and L. Collins, "A remote sensing approach to mapping fire severity in south-eastern australia using sentinel 2 and random forest," *Remote Sensing of Environment*, vol. 240, p. 13, 2020.

[6] Z. Xu, X. Su, and Y. Zhang, "Forest fire prediction based on support vector machine," *Chinese Agricultural Science Bulletin*, vol. 28, pp. 126–131, 2012.

[7] P. Cortez and A. Morais, "A data mining approach to predict forest fires using meteorological data." pp. 512–523, 2007.

[8] J. E. Halofsky, D. L. Peterson, and B. J. Harvey, "Changing wildfire, changing forests: the effects of climate change on fire regimes and vegetation in the pacific northwest, usa," *Fire Ecology*, vol. 16, no. 1, p. 4, 2020.

[9] M. Flannigan, B. Stocks, and B. Wotton, "Climate change and forest fires," *Science of The Total Environment*, vol. 262, pp. 221–229, 2000.

[10] D. J. Arrue BC, Ollero A, "An intelligent system for false alarm reduction in infrared forest-fire detection." *IEEE Intell Syst Appl*, vol. 3, pp. 64–73, 2000.

[11] G. Sakr, I. Elhajj, G. Mitri, and U. Wejinya, "Artificial intelligence for forest fire prediction," *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM*, pp. 1311–1316, 07 2010.

[12] D.-T. Bui, B. Pradhan, H. Nampak, Q.-T. Bui, Q-A.Tran, and Q.-P. Nguyen, "Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibilitgy modeling in a high-frequency tropical cyclone area using gis," *Journal of Hydrology*, vol. 540, pp. 317–330, 2016.

[13] R. S. Aakash, M. Nishanth, R. Rajageethan, R. Rao, and R. Ezhilarasie, "Data mining approach to predict forest fire using fog computing," *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1582–1587, 2018.

[14] P. Cortez and A. Morais, "A data mining approach to predict forest fires using meteorological data." pp. 512–523, 2007.

[15] "Predicting late-successional fire refugia pre-dating european settlement in the wenatchee mountains," *Forest Ecology and Management*, vol. 95, no. 1, pp. 63–77, 1997.

[16] H. reza Pourghasemi, M. Beheshtirad, and B. Pradhan, "A comparative assessment of prediction capabilities of modified analytical hierarchy process (m-ahp) and mamdani fuzzy logic models using netcad-gis for forest fire susceptibility mapping," *Geomatics, Natural Hazards and Risk*, vol. 7, no. 2, pp. 861–885, 2016.

[17] R. S. Bhadoria, M. K. Pandey, and P. Kundu, "Rvfr: Random vector forest regression model for integrated and enhanced approach in forest fires predictions," *Ecological Informatics*, vol. 66, p. 101471, 2021.

[18] R. S. Aakash, M. Nishanth, R. Rajageethan, R. Rao, and R. Ezhilarasie, "Data mining approach to predict forest fire using fog computing," *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1582–1587, 2018.

[19] Z. Wu, H. HS, J. Yang, Z. Liu, and Y.Liang., "Relative effects of climatic and local factors on fire occurrence in boreal forest landscapes of northeastern china," *Sci Total Environ*, vol. 493, pp. 472–80, 2014.

[20] A. Elshewey and A. Elsonbaty, "Forest fires detection using machine learning techniques," *Xi'an Jianzhu Keji Daxue Xuebao/Journal of Xi'an University of Architecture and Technology*, vol. XII, p. 2020, 10 2020.

[21] A. Nebot and F. Mugica, "Forest fire forecasting using fuzzy logic models," *Forests*, vol. 12.

[22] S. Al-Janabi, I. Al-Shourbaji, and M. A. Salman, "Assessing the suitability of soft computing approaches for forest fires prediction," *Applied Computing and Informatics*, vol. 14, no. 2, pp. 214–224, 2018.

[23] H. Liang, M. Zhang, and H. Wang, "A neural network model for wildfire scale prediction using meteorological factors," *IEEE Access*, vol. 7, pp. 1–1, 12 2019.

[24] Z. Guoli, W. Ming, and K. Liu, "Forest fire susceptibility modeling using a convolutional neural network for yunnan province of china," *International Journal of Disaster Risk Science*, vol. 10, pp. 386–406, 2019.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," vol. 25, 2012.

[26] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "Lstm fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2018.

[27] Z. Li, Y. Huang, X. Li, and L. Xu, "Wildland fire burned areas prediction using long short-term memory neural network with attention mechanism." vol. 57.

[28] R. Sharma, S. Rani, , and Memon, "Intelligence approaches smart approach for fire prediction under uncertain conditions using machine learning," *Multimed Tools Appl*, vol. 79, 2020.

[29] H. Hong, M. Panahi, A. Shirzadi, T. Ma, J. Liu, A. Zhu, W. Chen, I. Kougias, and N. Kazakis, "Flood susceptibility assessment in hengfeng area coupling adaptive neurofuzzy inference system with genetic algorithm and differential evolution," *Science of The Total Environment*, vol. 621, pp. 1124–1141, 2018.

[30] L. Batina, B. Gierlichs, E. Prouff, M. Rivain, F.-X. Standaert, and N. Veyrat-Charvillon, "Mutual information analysis: a comprehensive study," *Journal of Cryptology*, vol. 24, no. 2, pp. 269–291, Apr 2011. [Online]. Available: https://doi.org/10.1007/s00145-010-9084-8

[31] N. Arunraj, D. Ahrens, and M. Fernandes, "Application of sarimax model to forecast daily sales in food retail industry," *International Journal of Operations Research and Information Systems*, vol. 7, pp. 1–21, 04 2016.

[32] D. A. Wood, "Prediction and data mining of burned areas of forest fires: Optimized data matching and mining algorithm provides valuable insight," *Artificial Intelligence in Agriculture*, vol. 5, pp. 24–42, 2021.

Revolutionizing Historical Document Digitization: LSTM-Enhanced OCR for Arabic Handwritten Manuscripts

Safiullah Faizullah[1], Muhammad Sohaib Ayub[2], Turki Alghamdi[3], Toqeer Syed Ali[4],
Muhammad Asad Khan[5], Emad Nabil[6]
Faculty of Computer and Information Systems, Islamic University of Madinah, KSA[1,3,4,6]
Department of Computer Science, Lahore University of Management Sciences, Pakistan[2]
Department of Telecommunication, Hazara University, Pakistan[5]

***Abstract*—Optical Character Recognition (OCR) holds immense practical value in the realm of handwritten document analysis, given its widespread use in various human transactions. This scientific process enables the conversion of diverse documents or images into analyzable, editable, and searchable data. In this paper, we present a novel approach that combines transfer learning and Arabic OCR technology to digitize ancient handwritten scripts. Our method aims to preserve and enhance accessibility to extensive collections of historically significant materials, including fragile manuscripts and rare books. Through a comprehensive examination of the challenges encountered in digitizing Arabic handwritten texts, we propose a transfer learning-based framework that leverages pre-trained models to overcome the scarcity of labeled data for training OCR systems. The experimental results demonstrate a remarkable improvement in the recognition accuracy of Arabic handwritten texts, thereby offering a highly promising solution for the digitization of historical documents. Our work enables the digitization of large collections of ancient historical materials, including manuscripts and rare books characterized by delicate physical conditions. The proposed approach signifies a significant step towards preserving our cultural heritage and facilitating advanced research in historical document analysis.**

Keywords*—Optical character recognition; transfer learning; Arabic OCR; image processing; classification; convolutional neural network

I. Introduction

Around 1.8 billion people in the world speak the Arabic language. Arabic writing is unique and semi-cursive in both printed and handwritten forms. Arabic OCR systems are of two types, online and offline, aiming to convert Arabic text images into machine-readable words. Online systems use special equipment like a pen and tablet, while offline systems use scanners. There are open issues in Arabic OCR, such as generalization ability, the use of deep learning, lack of standard taxonomy, large-scale evaluation, and reproducible research [1], [2], [3].

The method used to process the documents or images to extract text is represented as OCR. These images or documents be in different forms, like scanned or digitized. This process helps extract text from these documents in an editable form for a machine to edit it. This process exists in two types, i.e. online and offline. In an online OCR system, real-time text recognition will be performed like writing on digital notebooks [4]. Whereas, written documents or images taken from some written sources are used in the offline OCR system [5]. As we focus on Arabic and dealing with handwritten documents or images, it comes in both offline and online types. Therefore, recognizing the text more accurately is the issue because the Arabic language has unique and challenging characteristics compared to other languages [6]. Its characters come in different shapes concerning their position in words, increasing the difficulty level for recognition, and when the document is of low quality or in different writing styles, fonts, cursive nature, and quality documents [7].

Some available tools help extract and recognize the Arabic text from the documents, like Tesseract [8], OCR Space [9], OmniPage [10], easy-OCR [11] and others. From all these, Tesseract gives a better result, but it needs more training and data to recognize handwritten Arabic text with high character, word, and overall text accuracy. By keeping this point in mind, we decided to use the transfer learning method on Tesseract to achieve a high recognition rate for offline handwritten Arabic text. It needs more training datasets, including ground truth, images, and box files. For this purpose, we write a script that makes images of the ground truth to fast forward the process and then makes the box files, and then these files are used together for transfer learning.

OCR can be used in many fields of life where it makes work easy, efficient, and digital. Hospitals process patient and insurance company files on the computer to make a digitized record as these kinds of records are handwritten, which is why OCR needs handwritten documents as well and uses them in many more fields [12], [13], [14].

A. Problem Formulation

Many OCR systems are available for different languages with different features, like unilingual or multilingual OCR systems. Much research is done in English and other Germanic languages, especially in noncursive scripts, because these are easier to be processed by the OCR systems. However, scripts like Arabic are challenging to recognize due to their cursive nature, the appearance of Arabic letters in different shapes in different words as shown in Table I, and diacritics that come above or below the Arabic word or letter. It changes the meaning of words as they are just minor signs, as shown in Table II. Arabic script is written from right to left, and some writing

styles change the appearance of the words, like in Fig. 1. Due to these complexities, Arabic OCR is a challenging research area that needs more consideration. Limited publicly available datasets exist for research purposes, and different techniques are available. However, each technique has pros and cons that limit character and word accuracy in the preprocessing and segmentation phases [15], [16].

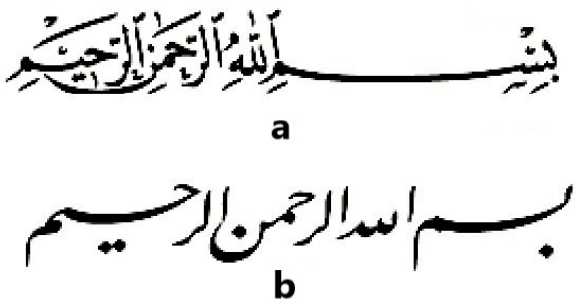


Fig. 1. Arabic sentence in two different writing styles, i.e. (a) Nask and (b) Nastaliq.

Table I. Arabic Letters and their Different Shapes based on the Position in a Word

Isolated Form	Initial Form	Middle Form	End Form
ا	ا	ـا	ـا
ب	بـ	ـبـ	ـب
ت	تـ	ـتـ	ـت
ث	ثـ	ـثـ	ـث
ج	جـ	ـجـ	ـج
ح	حـ	ـحـ	ـح
خ	خـ	ـخـ	ـخ
د	د	ـد	ـد
ذ	ذ	ـذ	ـذ
ر	ر	ـر	ـر
ز	ز	ـز	ـز
س	سـ	ـسـ	ـس
ش	شـ	ـشـ	ـش
ص	صـ	ـصـ	ـص
ض	ضـ	ـضـ	ـض
ط	طـ	ـطـ	ـط
ظ	ظـ	ـظـ	ـظ
ع	عـ	ـعـ	ـع
غ	غـ	ـغـ	ـغ
ف	فـ	ـفـ	ـف
ق	قـ	ـقـ	ـق
ک	کـ	ـکـ	ـک
ل	لـ	ـلـ	ـل
م	مـ	ـمـ	ـم
ن	نـ	ـنـ	ـن
ه	هـ	ـهـ	ـه
و	و	ـو	ـو
ي	يـ	ـيـ	ـي

B. Problem Motivation

The digitization of historical documents stands as a critical endeavor in the preservation and dissemination of our global cultural heritage. These documents, ranging from ancient manuscripts to letters and administrative records, encapsulate not only the factual history but also the intellectual, social, and cultural dynamics of past societies. Digitization offers a bulwark against the relentless march of time, safeguarding these irreplaceable insights from the ravages of physical degradation, environmental hazards, and the obscurity of inaccessibility. Moreover, it democratizes access to knowledge, transcending geographical and temporal barriers to make these treasures universally accessible, fostering a broader understanding and appreciation of human history. In the realm of Arabic historical documents, which are rich in linguistic and cultural nuances, digitization is not merely a technical challenge but a crucial step towards preserving a significant part of the world's intangible cultural heritage. Thus, our research is motivated by the imperative to advance Optical Character Recognition (OCR) technologies, specifically tailored to Arabic script, to facilitate the efficient and accurate digitization of these documents, ensuring their preservation and accessibility for future generations.

Table II. Position of Arabic Diacritics in the Script Changes the Meaning and Sound of the Words

Diacritics Type	Diacritics Shape
Fathah	◌َ
Kasrah	◌ِ
Dammah	◌ُ
Alif Khanjariyah	◌ٰ
Sukūn	◌ْ
Tanwin	◌ً ◌ٌ ◌ٍ
Shaddah	◌ّ

The motivation for this research paper on Arabic handwritten OCR stems from the challenges associated with recognizing and digitizing handwritten Arabic documents. While there are existing OCR systems for recognizing printed Arabic text, there has been less research on developing accurate models for recognizing handwritten Arabic characters and numbers. This is problematic because handwritten documents are important cultural artifacts that need to be preserved, but they are at risk of being lost if not digitized. Furthermore, the lack of publicly available datasets for Arabic handwritten OCR makes it difficult for researchers to develop and evaluate new models. Therefore, this paper aims to contribute to the field of Arabic handwritten OCR by presenting a new dataset and an accurate OCR model that can be used to preserve and make accessible handwritten Arabic documents.

We make the following contributions to significantly advance the state-of-the-art in Optical Character Recognition (OCR) for Arabic handwritten texts:

1) We present an enhanced OCR accuracy for Arabic handwritten texts through the novel application of transfer learning techniques to the Tesseract OCR engine, substantially reducing common recognition errors.
2) Our work includes the compilation and preparation of a comprehensive dataset, which comprises high-resolution `.tiff` images, `.box` files, `ground truth` files, and `dictionary files`. This dataset not only supports our model's training but also serves as a resource for the broader research community.
3) We introduce a robust framework for evaluating OCR accuracy, utilizing Character Error Rate (CER) and Word Error Rate (WER) as the principal metrics. This framework facilitates a thorough quantitative analysis and validation of our approach.
4) Through detailed **visual demonstrations of our**

model's efficacy, we provide clear evidence of our methodology's effectiveness and its practical implications for Arabic text digitization.

These contributions underscore the novelty and significance of our work, showcasing its potential to significantly impact the OCR field by enhancing the recognition accuracy of Arabic handwritten texts and promoting the digitization and preservation of historical documents.

II. Related Work

The exploration of Optical Character Recognition (OCR) technologies, especially in the realm of Arabic handwritten texts, reveals a landscape marked by both advancements and persistent challenges. While the strides in OCR methodologies have paved the way for notable achievements, the unique intricacies of Arabic script—ranging from its cursive nature to the prevalence of diacritics—pose specific hurdles that remain inadequately addressed. This gap not only underscores the necessity for innovative approaches but also serves as the cornerstone of our motivation. Our investigation into the related work illuminates the breadth of strategies previously employed, yet simultaneously highlights a critical void in the application of advanced machine learning techniques, such as transfer learning, to the nuances of Arabic handwriting recognition. It is this intersection of opportunity and challenge that our research aims to navigate, propelled by the conviction that enhancing OCR accuracy for Arabic texts not only contributes to the technological domain but also fosters the preservation and accessibility of cultural heritage. As such, our work seeks not only to bridge the existing gaps identified through our review but to set a new benchmark for accuracy and efficiency in the recognition of Arabic handwritten documents.

The Arabic language has complexities for developing OCR systems due to its cursive nature and morphological structure. Applied approaches and methods in this area are compared and show the best-performing approach. This study shows that the hidden Markov model (HMM) gives an accuracy of 95.6%. As the survey is working on handwritten characters recognition of Arabic, it shows the recognition by an automatic capitalization of OCR using the hidden Markov model [17].

Arabic text recognition system faces many challenges due to its cursive nature, different shapes of characters, diacritics, and writing styles [7], [18]. Handwritten text recognition comes in two formats, i.e. document (manuscripts) and online (tablets) [19]. A proper pipeline is required to overcome the challenges and get a high recognition rate [20]. Urdu and Arabic are the same, i.e. Urdu uses all of the Arabic characters. Recent advances discuss state-of-the-art research about Urdu and Arabic Naskh and Nastaliq scripts. It has also discussed the dataset and its different forms, i.e. printed, scanned, and handwritten, and the pipeline used to process the data, i.e. preprocessing, segmentation, classification, recognition, and post-processing to get better text recognition [21], [22], [23].

As mentioned above, most researchers also work on Arabic and use transfer learning. However, they only concentrated on characters and numbers, which are easier to recognize through transfer learning on models like Alex-Net and Google-Net. Moreover, they are already trained on the image of different objects and give high recognition on a single letter or number images [24].

The authors of [25] discuss the challenges of handwritten Arabic text classification and recognition and evaluate the different deep learning models, i.e. ResNet50, ResNet101, VGG16, VGG19, AlexNet, GoogleNet, and ResNet18 using transfer learning techniques. To overcome the challenges, use handwritten images of text written by a native or non-native person. The dataset consists of 22 subjects equally written by native and non-native writers. After using different models, results show that GoogleNet is the model that achieves 93.2% accuracy on the native dataset and 95.2% on the non-native dataset.

Sahlol et al. [26] describe a hybrid machine-learning approach for handwritten Arabic character recognition using optical character recognition (OCR) systems. This approach combines neighborhood rough sets with a binary whale optimization algorithm for feature selection. The proposed method outperformed state-of-the-art and deep neural networks regarding recognition rate and computational time. However, some misclassified failure cases occurred due to the context of appearance.

Authors of [27] present a new model for Arabic document information retrieval (ADIR) using OCR services. They used datasets written by 60 writers containing 16,800 Arabic letters and applied them to preprocess techniques such as binarization, noise removal, skew correction, and page and zone segmentation. They discussed the challenges of segmentation and recognition for Arabic script, particularly with cursive writing and compound graphemes. The Arabic document information retrieval (ADIR) system achieved a classification success rate of 100% for test images. The paper also describes the service description for ADIR, which includes the user interface (UI) and a server address for communication between clients and services.

Rahal et al. [28] discussed the difficulties of automatic Arabic text recognition due to the language's cursive nature, character similarities, large vocabulary, and the use of multiple font sizes. This paper proposed a novel hybrid network that combines Hidden Markov Models (HMMs) and a Bag-of-Feature (BoF) framework, which is based on a deep Sparse Auto-Encoder (SAE) for feature extraction. The system eliminates the need for preprocessing stages like baseline estimation and slant normalization. Instead, the SAE automatically selects the best weights for visual words for each local descriptor while concurrently learning the best dictionary, making it suitable for irregular, variable-size, mixed-font, high- and low-resolution documents. The system was tested on four different datasets and performed well on each one.

Zanona et al. [29] present a model for recognizing handwritten Arabic characters that use preprocessing functions and contour analysis to produce a vector for recognition by a neural network. The system was tested on private data and achieved 98% accuracy on the complete dataset and 99.4% precision. The classification system uses a segmentation operation [30] and a multilayer feed-forward neural network (FFNN). In another research, an Arabic handwritten text recognition system was designed that extracts and recognizes single-line text and converts that extracted or predicted text into individual words

and their characters, achieving an 83% recognition rate.

Many advanced systems, like Mathpix, Digital Ink API, ML Kit Text, Read-Ink, MyScript, GoodNotes, and Mazec, were developed in the last decade to resolve challenges in handwritten text recognition. Multiple methods and techniques are used for this purpose, like dynamic time warping, hidden Markov models, and artificial neural networks, which use a pipeline to process the data correctly. These methods use the data and enhance the recognition rate, which shows impressive results [31].

Some systems are used for the Arabic text OCR, i.e. Tesseract, Textract, and Document AI. These OCR systems work for different languages, like English and Arabic. In this research, historical documents are used, and these documents are multilingual and contain English and Arabic text. The English dataset contains historical text from books that scan with various fonts. In contrast, the Arabic dataset contains articles from online resources in a single widely used font. The author found that Textract and Tesseract performed slightly better on gray-scale test images than on color versions, but Tesseract was more sensitive to noise than the other two engines. It shows that Document AI and Textract give better results than Tesseract. However, Document AI and Textract have higher noise in their images even after applying noise removal technique in it [32], which recommends self-training or transfer learning to the Tesseract, it will be able to give better results than others.

Most of the researchers work in Arabic handwritten OCR. However, they mostly used characters, digits, or other individuals in their experiments and showed better results as shown in Table III.

Authors in [35] proposed a technique based on text area detection and text recognition using pre-trained OCR systems, i.e. Tesseract, KerasOCR, and EasyOCR. This system deals with engineering documents, and a high recognition rate is significant. For this purpose, transfer learning is used, which helps to increase the word recognition rate and increase the overall text recognition rate as well.

III. Methodology

This section describes the methodology of our proposed work. We provide a descriptive analysis of our datasets. We have explained our transfer learning approach and each step of the experimentation.

A. Dataset

As datasets play an important role in training models, a more extensive dataset helps train the model more accurately, and the model can learn quickly. For deep and machine learning models, larger datasets are required, but due to handwritten text, data preparation is a tough job as first need to generate round truth. Handwritten text against ground truth and its image and box files generation to pass the model for text extraction, recognition, and further processing. Due to these challenges, the availability of a public dataset is very low. Moreover, every model requires a large amount of data to do its excellent training; by considering this problem, we moved toward transfer learning, and we know that for transfer learning models, a larger dataset is not an issue, and this technique helps the model to get better training even on smaller datasets.

Urdu and Arabic have many similarities, i.e. both are written from right to left, all of the Arabic characters are used in Arabic, but Urdu has just extra characters, their digits have the same shapes, their writing styles are also common, and Urdu borrows a large amount of vocabulary from Arabic that is about 30%.

Due to these similarities, the Urdu dataset can also be used for Arabic text. Bhatti et al. [36] presents an Urdu dataset based on the Urdu handwritten digits dataset and experiments using deep learning techniques and achieving a high accuracy rate.

An Urdu handwritten dataset is proposed by [37], which is based on Nastaliq handwritten text, i.e. UNHD, and used a bidirectional LSTM classifier and this dataset is written by 500 writers on A4 size paper as shown in Table IV that will be available on request. The address of the dataset link is attached [1].

Mostafa et al. [38] proposed a dataset that is based on text images shown in Fig. 2 with their ground truth as well. The dataset contains 270 million words and 1.6 billion characters. 12 fonts are used in the dataset, and the used text font size is 13.

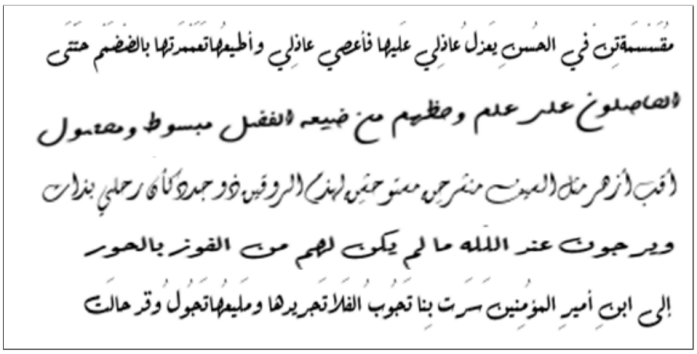


Fig. 2. Dataset sample of arabic text image that contains text with diacritics.

The authors of [39] proposed a handwritten dataset that is taken from different handwritten documents. A description of the dataset is shown in Table V and a sample image of the dataset is shown in Fig. 3.

Dataset selection:

For dataset selection, we find the Arabic text and then write a Python script that tests a script that gives the relative frequency of characters present in the text, and then we compare the relative frequency to that relative frequency [40], [41], as shown in the Table VI to avoid biases, this frequency distribution helps to use the characters in same quantity and position as they are used in normal text writing. Also, we test the script on each file separately to check their frequency. We then gave the text files to different people to write the handwritten text, and then we scanned that text and evaluated the handwritten dataset.

[1] https://sites.google.com/site/researchonurdulanguage1/databases

Table III. Previous Arabic Handwritten Text Research Focused on Character Recognition, Digit Recognition, Hijja Recognition, and Word Recognition using Image-based Data for Model Training

Paper	Model	Dataset	Accuracy
Albhattah [33]	Hybrid CNN with finetune	AHCD (characters and Hijja)	98%
	CNN with finetune	AHCD (characters)	92.4%
Alghyaline [34]	Deep CNN	APTI (words)	76.30%
	CNN-RNN	Alif dataset (words)	85.98%
	Deep CNN	HMBD (characters)	92.88%

Table IV. Description of the Urdu Nastaliq Handwritten Dataset, adaptable for Arabic Due to its Derivation from the Arabic Script

Urdu Nastaliq Handwritten Dataset	Description
Writers	500
Text lines	10,000
Words	312,000
Characters	1,872,000
Words written by a writer	624

Table V. Description of the Handwritten Dataset Extracted from Handwritten Documents

No. of Pages	No. of Lines	No. of Words	No. of Chars
1,000	18,000	35,000	252,000

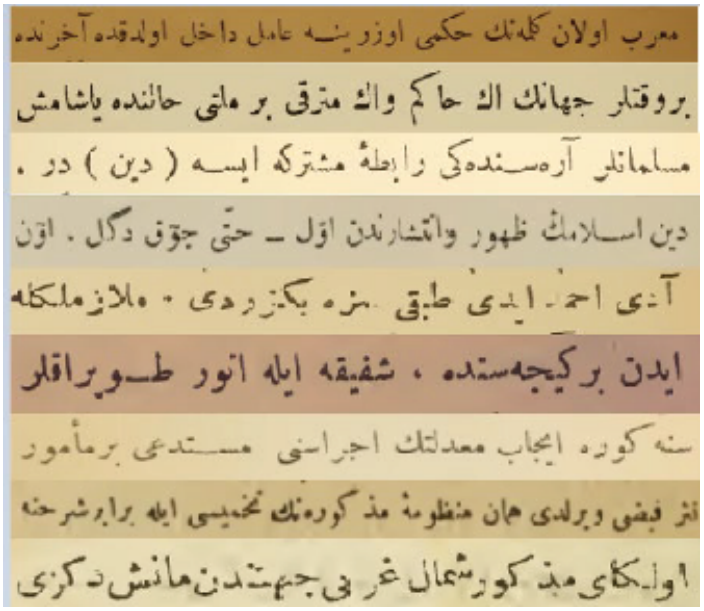


Fig. 3. Sample image of the handwritten dataset that is taken from the various Arabic handwritten documents.

1) Publicly Available Datasets: Some handwritten datasets are reported [2], but those datasets contain images of words because most researchers work on individual character and number recognition of Arabic. They used these datasets for pre-trained object image recognition models like ImageNet and AlexNet [19]. A description of these kinds of datasets is shown in Table VII.

Arabic MNIST dataset is also available, but it is also for individual characters and numbers. This database is split into training and testing datasets for Arabic characters and numbers, consisting of 60,000 images for training and 10,000 images for testing purposes of characters and the same for numbers; this dataset is available [3]. Furthermore, another dataset is available that is written by the contribution of 60 writers and contains 168,00 character images, and these images were scanned on 300dpi resolution [42].

Table VI. Comparison of Arabic Characters Frequency in the Selected Text and Overall Frequency of Arabic Characters

Characters	Frequency in selected text	Frequency in Arabic Language
ا	15.00%	14.61%
ل	14.13%	11.64%
م	6.56%	6.49%
ي	6.38%	7.25%
و	5.64%	5.40%
ن	5.41%	4.76%
ت	4.39%	4.58%
ع	4.15%	3.27%
ر	3.54%	4.53%
ب	2.70%	3.38%

Table VII. Description of Handwritten Images of Arabic Words Dataset used in Pre-trained Models

Statistics	AlexU-W	IFN/ENIT
Images	25,114	32,492
Training images	20,114	26,459
Testing images	5,029	6,033
Unique words	109	937
Maximum PAWS	3	10

A gold-standard dataset is available on GitHub [4], which contains Arabic books, their ground truth, images, and box files. That is much enough to train any model, text images based on one-line text, and a description of the gold standard set is shown in Table VIII. Some publicly available datasets of Arabic handwritten text, words, characters, and digits are also available, as shown in Table IX.

Table VIII. EDA of Arabic Gold-Standard Dataset Derived from Literature Books

Book	Pages	Lines	Words	Chars
IbnFaqihHamadhani.al-Buldan	79	1466	16909	92730
IbnAthir.al-kamil	40	794	12818	58481
Ibn Qutayba.Adab al-katib	55	794	7848	42230
al-Jahiz.al-Hayawan	65	992	11870	59191
al-Yacqubi.al-Tarikh	68	1050	13487	66341
al-Dhahabi.Tarikh al-islam	50	1110	11045	55047
Ibn al-Jawzi.al-Muntazam	50	938	13156	62574

We used a publicly available dataset for our experiments, which consists of images of handwritten text, ancient Arabic dataset, and printed Arabic text data for comparison and its ground truth as well [52], [53]. The dataset consists of about

[2]http://www.eng.alexu.edu.eg/~mehussein/alexu-word/

[3]http://yann.lecun.com/exdb/mnist/

[4]https://github.com/OpenArabic/OCR_GS_Data/tree/master/ara

Table IX. Publicly Available Datasets Focus Primarily on Arabic Ligatures and Digits

Dataset	Type of content
IFN/ENIT [43]	Handwritten Words
HACDB [44]	Handwritten Characters
KHATT [45]	Handwritten Text lines
SmartATID [46]	Printed & Handwritten Pages
Degraded historical [47]	Handwritten documents
Numeral [48]	Handwritten Digits
AHDBIFTR [49]	Handwritten images
ARABASE [50]	Handwritten Text
CENPARMI [51]	Handwritten subwords & digits

25,000 entries, which contain ground truth, images, box, and "lstmf" files, which is a complete set of the required dataset for training the Tesseract transfer learning, and these images consist of one-line text, and in a raw format that will be preprocessed in the preprocessing stage of Tesseract pipeline, as shown in Table X.

Table X. EDA of Datasets used for Evaluating Handwritten Arabic Text in Our Experiments

Dataset	Words	Char	Lines	Digits	Punc.
Printed lines	3,592	19,802	307	29	370
Words and lines	71,365	374,516	14,606	5,592	6,451
IbnFaqihHamadhani	104,845	577,392	15,296	1,713	9,948
Ancient Arabic	700	3,160	100	-	-

B. Transfer Learning using Tesseract

Tesseract is an open-source engine for Optical Character Recognition used to recognize text from images. It consists of the following steps to process the image from its raw shape to make it able to be used for OCR, i.e. preprocessing, Converting Image to Box file, Converting Box files to "lstmf" file.

While using the Tesseract base model, firstly, we passed the handwritten image to the model to check its accuracy and try to find the problem in the base model. After passing images in PNG or JPG, we write a script that takes all the images from the folder individually, passes them from the model, extracts text from the images, and saves them in a text file in another folder. After this, we pass both ground truth and extracted or predicted text files and find accuracy. For evaluation, we write a script using Levenshtein distance to find the character error rate (CER) and word error rate (WER) and then find the average CER and WER and their accuracy as the base model is trained on editable or computer-typed text. Therefore, we prepare the data for model [5] some dictionary files, the dataset contains ".tiff" files and ground truth and passed this data to the Tesseract for transfer learning purpose and consider starting point to the base model. After transfer learning, repeat the test for evaluation and get the results with more accuracy, CER and WER, as shown in Fig. 4.

The methodology depicted in Fig. 4 commences with the meticulous preparation of training data, a pivotal phase for the effective employment of transfer learning on the Tesseract OCR engine. This stage encompasses the assembly of high-resolution `.tiff` images of Arabic handwritten texts, alongside their corresponding `.box` files which delineate bounding boxes around each character, and `ground truth` files containing the verbatim text. Furthermore, `dictionary files` are curated to guide the model towards recognizing the expected lexical items. Subsequent to the data preparation, the Tesseract base model undergoes augmentation through transfer learning, enabling it to adapt to the nuances of Arabic handwritten text. This adaptation is facilitated by training the model with the prepared dataset, thereby enhancing its capability to accurately recognize Arabic characters and words. The retrained model is then evaluated against a set of unseen data to ascertain its accuracy, employing metrics such as the Character Error Rate (CER) and Word Error Rate (WER) derived through Levenshtein distance calculations. These metrics serve as indicators of the model's proficiency in character and word recognition, with lower CER and WER values signifying superior performance. This comprehensive approach underscores the efficacy of transfer learning in refining the Tesseract model's recognition accuracy for Arabic handwritten texts.

Preprocessing: In this step, Tesseract takes an image for preprocessing and improves the image quality by resizing the image into a standard size, converting the image into gray-scale, and applying filters to remove noise from the image like thresholding, erosion, blur.

Converting image into Box file: In this step, it converts the preprocessed image into a box file with plain text containing coordinates of the bounding boxes around each character or word of text present in an image. The following command is used to make a box file against the image. "Tesseract image.png output -l ara makebox". In this command, "image.png" is an image that needs to be recognized, and 'output' is the file that gets the recognized text and saves it, while "ara" is for language; here, we are using Arabic, and we used the keyword for it, same "eng" is for the English language.

Converting box file to "lstmf": This step takes the box file created in the previous step and makes a binary file containing the training data for Tesseract. This process contains the following steps like "tesseract image.png output -l ara box.train", Which will create "lstmf" files with the name of "ara.traineddata" in the same directory.

Evaluation: For evaluating, we need to compare the recognized text with the ground truth to get accuracy. For this purpose, we can use the following steps to get a better evaluation, i.e. CER, WER, and Overall accuracy. A higher accuracy, lower CER, and BER represent good results. CER gives the percentage of incorrectly recognized characters. It can be calculated simply by dividing the number of incorrect characters by the number of characters in the ground truth, as shown in Eq. (1) below.

$$\text{CER} = \frac{\text{Number of incorrect characters}}{\text{Total number of characters}} \tag{1}$$

WER gives the percentage of incorrectly recognized words. It can be calculated simply by dividing the number of incorrect words by the total number of words in the ground truth, as shown in Eq. (2) below.

$$\text{WER} = \frac{\text{Number of incorrect words}}{\text{Total number of words}} \tag{2}$$

[5] https://github.com/Shreeshrii/tesstrain-JSTORArabic/tree/master/data/

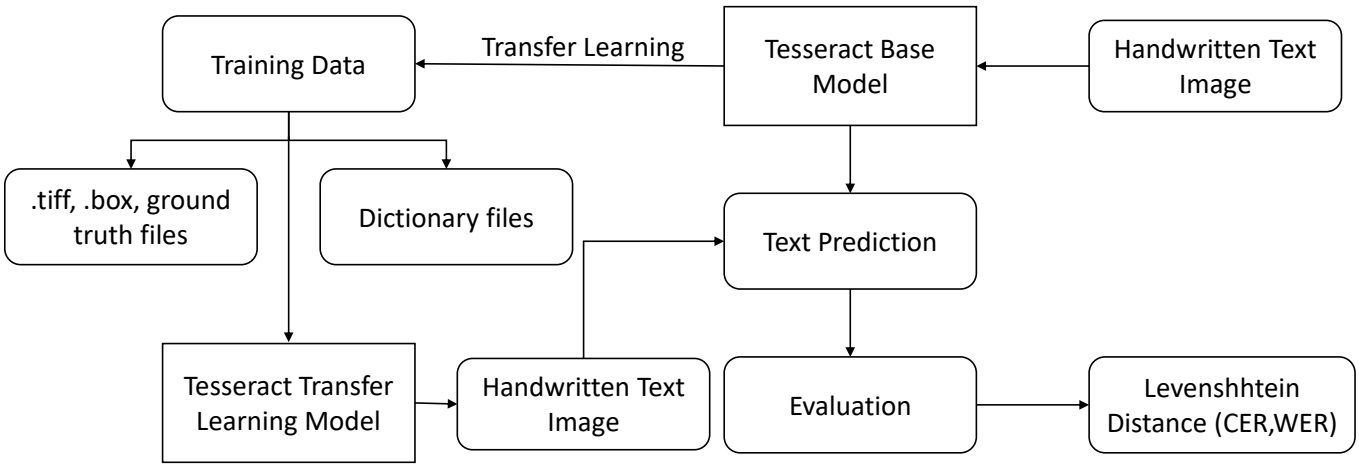


Fig. 4. Tesseract base-model, transfer learning process and evaluation methods.

Overall accuracy gives the percentage of correctly recognized words and characters. It can be calculated by dividing the number of correct words and characters by the total number of words and characters in the ground truth, as shown in Eq. (3) below.

$$\text{Accuracy} = \frac{\text{No. of correct } \{\text{words + characters}\}}{\text{Total no. of } \{\text{words + characters}\}} \quad (3)$$

IV. Results

The experiments were conducted on a Windows Dell laptop with an Intel(R) Core(TM) i7-6600U CPU @ 2.60GHz, 16GB DDR4 RAM, 512GB SSD storage, and an Intel® HD Graphics 520. The laptop ran Windows 10 and was connected to a stable power source throughout the experiments. The experiments were conducted in a controlled environment to minimize external factors that could affect the results. Additionally, Ubuntu 22.04.1 LTS was installed using WSL to run the Tesseract experiment.

In the next section, evaluation metrics, i.e. CER and WER, are discussed, and the overall accuracy of the predicted text against ground truth is described with the results. Then, a discussion of the results and experiments is presented.

A. Evaluation Metrics

After transfer learning of Tesseract, we evaluate the model. For this purpose, firstly, give handwritten text image files to the model that generates box and "lstmf" file and then gives the predicted text of all images. Then, this predicted text compares with ground truth, which is also available in a text file. As mentioned above, the dataset is split into 80% for the training and 20% for testing. Furthermore, we used two types of evaluation, i.e. our evaluation that gives us the overall accuracy of the experiment, CER and WER. For more detail, we use an open-source evaluation tool named "OCREVALUATION", which compares ground truth and the predicted text that elaborates more openly, as shown in Fig. 5. This sample image has two sections: one represents the ground truth on the left, and the other is about the predicted text. The ground truth section represents the characters or words by different colors that are predicted wrong in the predicted text section.

The base model gives an overall accuracy of 23.30%, an average CER of 31.57%, and an average WER of 65.95% on handwritten text images of the Arabic language. However, after transfer learning, it gives an overall accuracy of 87.89% and gives the average CER of 14.02% and average WER of 41.39%, which is relatively better, also shown in Table XI. It also depends on the size of the dataset, and the total time taken by the training is 21 hours. Then, evaluate the second dataset, which contains 5526 images of text that are also based on 1 line text. For these images, we also have ground truth for evaluation purposes and then generate "lstmf" and box files for each image and the ground truth file. After evaluation, we get a character error rate (CER) of 14.85%, a word error rate (WER) of 40.30%, and achieve an overall accuracy of 85.53% also shown in Table XI.

Table XI. Comparison of Tesseract base Model with our Transfer Learning-based Model for Arabic Handwritten Text Recognition

Model	CER	WER	Accuracy%
Tesseract Base Model	31.57%	65.95%	23.30%
Transfer Learning (Dataset 1)	14.02%	41.39%	87.89%
Transfer Learning (Dataset 2)	14.85%	40.30%	85.53%

To check the accuracy of the printed text of this retrained model, we test a dataset that contains a total of 6118 files, which are divided into image and ground truth files, and that took about 5 hours to evaluate. It achieves an accuracy of 94.94%, character error rate, and word error rate based on order dependent and independent are shown in Table XII.

Table XII. Arabic Printed Images Text Recognition with Order Dependent and Order Independent

Features	Stats
Number of files	3,059
CER (order independent)	5.62%
WER(order independent)	19.08%
CER (order dependent)	5.55%
WER(order dependent)	17.85%

Fig. 5. Spotting differences between ground truth (gt.txt) and predicted (ara.OCR.txt) arabic handwritten text.

B. Results and Discussion

For this problem, we have chosen the open-source OCR engine Tesseract as the base model and applied the transfer learning approach to get better Arabic handwritten text recognition results. Firstly, we tested the base model to know its results on Arabic text; we decided to evaluate the computer-generated Arabic text first. After evaluation, we found an average CER of 14.02% and an average WER of 41.39% and got an overall accuracy of about 87.89%. Then, we tested the handwritten images on the base model and got bad results, giving an average CER of 31.57%, WER of 65.95%, and accuracy of 23.30%.

Some challenges while recognizing text are that Arabic handwriting can vary significantly between individuals, making it difficult for Tesseract to recognize characters accurately. This variation can be due to factors such as writing style, speed of writing, and individual handwriting quirks. Arabic script includes diacritical marks, which are symbols that indicate vowel sounds. These marks can be challenging for Tesseract to recognize accurately, especially when small or poorly written.

Handwritten text may contain noise or distortion due to uneven ink distribution, smudging, or poor image quality. This can make it difficult for Tesseract to recognize characters accurately. Handwritten text may be oriented or aligned in various ways, making it challenging for Tesseract to recognize the correct characters or words. Additionally, training Tesseract with a more extensive and diverse dataset of Arabic handwriting may improve its accuracy.

After this, we apply the transfer learning approach and train the model by giving handwritten images and their ground truth. After 21 hours of training, the transfer learning gives outstanding results, with CER being 14.02%, WER being 41.39%, and overall accuracy of 87.89%, which is impressive. Then, evaluate the second dataset, which contains 5526 images of text that are also based on 1 line text. After evaluation, we get a character error rate (CER) of 14.85%, a word error rate (WER) of 40.30%, and achieve an overall accuracy of 85.53%.

V. Conclusion

In conclusion, our paper has presented an effective OCR method for handwritten Arabic text recognition using a transfer learning approach with Tesseract. Our evaluation results show a significant improvement in word and character recognition accuracy compared to previous models. The applied transfer learning technique achieved an average CER of 14.02% and an average WER of 41.39% and got an overall accuracy of about 87.89%. Then we decided to test the handwritten images on the base model to get the comparison result with the retrained model by using the transfer learning technique, giving an average CER of 31.57%, WER of 65.95%, and accuracy of 23.30%. These findings suggest that transfer learning can be a valuable technique for improving OCR accuracy in challenging languages such as Arabic and may provide a promising direction for future research in this field. Overall, our work demonstrates the potential of leveraging existing knowledge and models to improve the performance of OCR systems.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

Conceptualization: S.F., M.S.A., T.S.A., M.A.K. and A.A.; Data curation: M.S.A. and T.S.A.; Formal analysis: S.F.; Funding acquisition: S.F. and A.A.; Investigation: M.S.A., T.S.A. and M.A.K.; Methodology: M.S.A. and T.S.A.; Project administration: S.F. and M.A.K.; Resources: S.F. and A.A.; Software: M.S.A. and T.S.A.; Supervision: S.F. and M.A.K.; Validation: A.A.; Visualization: M.S.A. and T.S.A.; Writing – original draft: S.F., M.S.A., T.S.A., M.A.K. and A.A.; Writing - review & editing: S.F., M.S.A., T.S.A., M.A.K. and A.A.;

Acknowledgments

This work is funded by the Deputyship of Research & Innovation, Ministry of Education in Saudi Arabia, through project number 964. In addition, the authors would like to express their appreciation for the support provided by the Islamic University of Madinah.

References

[1] S. Djaghbellou, A. Bouziane, A. Attia, and Z. Akhtar, "A survey on arabic handwritten script recognition systems," *International Journal of Artificial Intelligence and Machine Learning (IJAIML)*, vol. 11, no. 2, pp. 1–17, 2021.

[2] L. S. Al-Homed, K. M. Jambi, and H. M. Al-Barhamtoshy, "A deep learning approach for arabic manuscripts classification," *Sensors*, vol. 23, no. 19, p. 8133, 2023.

[3] R. Najam and S. Faizullah, "Analysis of recent deep learning techniques for arabic handwritten-text ocr and post-ocr correction," *Applied Sciences*, vol. 13, no. 13, p. 7568, 2023.

[4] B.-G. Han, J. T. Lee, K.-T. Lim, and D.-H. Choi, "License plate image generation using generative adversarial networks for end-to-end license plate character recognition from a small set of real images," *Applied Sciences*, vol. 10, no. 8, p. 2780, 2020.

[5] A. F. d. S. Neto, B. L. D. Bezerra, and A. H. Toselli, "Towards the natural language processing as spelling correction for offline handwritten text recognition systems," *Applied Sciences*, vol. 10, no. 21, p. 7711, 2020.

[6] K. M. Nahar, I. Alsmadi, R. E. Al Mamlook, A. Nasayreh, H. Gharaibeh, A. S. Almuflih, and F. Alasim, "Recognition of arabic air-written letters: Machine learning, convolutional neural networks, and optical character recognition (ocr) techniques," *Sensors*, vol. 23, no. 23, p. 9475, 2023.

[7] S. Faizullah, M. S. Ayub, S. Hussain, and M. A. Khan, "A survey of ocr in arabic language: Applications, techniques, and challenges," *Applied Sciences*, vol. 13, no. 7, p. 4584, 2023.

[8] R. Smith, "An overview of the tesseract ocr engine," in *Ninth international conference on document analysis and recognition (ICDAR 2007)*, vol. 2. IEEE, 2007, pp. 629–633.

[9] T. C. Wei, U. Sheikh, and A. A.-H. Ab Rahman, "Improved optical character recognition with deep neural network," in *2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*. IEEE, 2018, pp. 245–249.

[10] F. Harbuzariu, C. Irimia, and A. Iftene, "Official document text extraction using templates and optical character recognition," in *2023 International Conference on Innovations in Intelligent Systems and Applications (INISTA)*. IEEE, 2023, pp. 1–4.

[11] N. Awalgaonkar, P. Bartakke, and R. Chaugule, "Automatic license plate recognition system using ssd," in *2021 International Symposium of Asian Control Association on Intelligent Robotics and Industrial Automation (IRIA)*. IEEE, 2021, pp. 394–399.

[12] A. Kumar, P. Singh, and K. Lata, "Comparative study of different optical character recognition models on handwritten and printed medical reports," in *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*. IEEE, 2023, pp. 581–586.

[13] F. Azzam, M. Jaber, A. Saies, T. Kirresh, R. Awadallah, A. Karakra, H. Barghouthi, and S. Amarneh, "The use of blockchain technology and ocr in e-government for document management: Inbound invoice management as an example," *Applied Sciences*, vol. 13, no. 14, p. 8463, 2023.

[14] H. Butt, M. R. Raza, M. J. Ramzan, M. J. Ali, and M. Haris, "Attention-based cnn-rnn arabic text recognition from natural scene images," *Forecasting*, vol. 3, no. 3, pp. 520–540, 2021.

[15] S. Bergamaschi, S. De Nardis, R. Martoglia, F. Ruozzi, L. Sala, M. Vanzini, and R. A. Vigliermo, "Novel perspectives for the management of multilingual and multialphabetic heritages through automatic knowledge extraction: The digitalmaktaba approach," *Sensors*, vol. 22, no. 11, p. 3995, 2022.

[16] F. M. Nashwan, M. A. Rashwan, H. M. Al-Barhamtoshy, S. M. Abdou, and A. M. Moussa, "A holistic technique for an arabic ocr system," *Journal of Imaging*, vol. 4, no. 1, p. 6, 2017.

[17] A. S. Shaker, "A survey for an automatic transliteration of arabic handwritten script," *Journal of Physics: Conference Series*, vol. 1530, no. 1, p. 012094, 2020.

[18] P. Ahmed and Y. Al-Ohali, "Arabic character recognition: Progress and challenges," *Journal of King Saud University-Computer and Information Sciences*, vol. 12, pp. 85–116, 2000.

[19] M. Awni, M. I. Khalil, and H. M. Abbas, "Offline Arabic handwritten word recognition: A transfer learning approach," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 10, pp. 9654–9661, 2022.

[20] S. Naz, K. Hayat, M. I. Razzak, M. W. Anwar, S. A. Madani, and S. U. Khan, "The optical character recognition of urdu-like cursive scripts," *Pattern Recognition*, vol. 47, no. 3, pp. 1229–1248, 2014.

[21] A. M. Alshanqiti, S. Albouq, A. B. Alkhodre, A. Namoun, and E. Nabil, "Employing a multilingual transformer model for segmenting unpunctuated arabic text," *Applied Sciences*, vol. 12, no. 20, p. 10559, 2022.

[22] A. Qaroush, B. Jaber, K. Mohammad, M. Washaha, E. Maali, and N. Nayef, "An efficient, font independent word and character segmentation algorithm for printed arabic text," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 1, pp. 1330–1344, 2022.

[23] A. Alshanqiti, A. Namoun, A. Alsughayyir, A. M. Mashraqi, A. R. Gilal, and S. S. Albouq, "Leveraging distilbert for summarizing arabic text: an extractive dual-stage approach," *IEEE Access*, vol. 9, pp. 135 594–135 607, 2021.

[24] M. A. KO and S. Poruran, "OCR-nets: variants of pre-trained CNN for Urdu handwritten character recognition via transfer learning," *Procedia Computer Science*, vol. 171, pp. 2294–2301, 2020.

[25] A. A. Almisreb, S. Turaev, M. A. Saleh, S. A. M. Al Junid *et al.*, "Arabic Handwriting Classification using Deep Transfer Learning Techniques," *Pertanika Journal of Science & Technology*, vol. 30, no. 1, pp. 641–654, 2022.

[26] A. T. Sahlol, M. Abd Elaziz, M. A. Al-Qaness, and S. Kim, "Handwritten arabic optical character recognition approach based on hybrid whale optimization algorithm with neighborhood rough set," *IEEE Access*, vol. 8, pp. 23 011–23 021, 2020.

[27] H. M. Al-Barhamtoshy, K. M. Jambi, S. M. Abdou, and M. A. Rashwan, "Arabic documents information retrieval for printed, handwritten, and calligraphy image," *IEEE Access*, vol. 9, pp. 51 242–51 257, 2021.

[28] N. Rahal, M. Tounsi, A. Hussain, and A. M. Alimi, "Deep sparse auto-encoder features learning for arabic text recognition," *IEEE Access*, vol. 9, pp. 18 569–18 584, 2021.

[29] M. A. Zanona, A. Abuhamdah, and B. M. El-Zaghmouri, "Arabic hand written character recognition based on contour matching and neural network." *Comput. Inf. Sci.*, vol. 12, no. 2, pp. 126–137, 2019.

[30] A. Zoizou, A. Zarghili, and I. Chaker, "A new hybrid method for arabic multi-font text segmentation, and a reference corpus construction," *Journal of King Saud University-Computer and Information Sciences*, vol. 32, no. 5, pp. 576–582, 2020.

[31] T. Ghosh, S. Sen, S. M. Obaidullah, K. Santosh, K. Roy, and U. Pal, "Advances in online handwritten recognition in the last decades," *Computer Science Review*, vol. 46, p. 100515, 2022.

[32] T. Hegghammer, "OCR with Tesseract, Amazon Textract, and Google Document AI: a benchmarking experiment," *Journal of Computational Social Science*, vol. 5, no. 1, pp. 861–882, 2022.

[33] W. Albattah and S. Albahli, "Intelligent Arabic Handwriting Recognition Using Different Standalone and Hybrid CNN Architectures," *Applied Sciences*, vol. 12, no. 19, p. 10155, 2022.

[34] S. Alghyaline, "A Printed Arabic Optical Character Recognition System using Deep Learning," *Journal of Computer Science*, vol. 18, no. 11, pp. 1038–1050, 2022.

[35] W. Khallouli, R. Pamie-George, S. Kovacic, A. Sousa-Poza, M. Canan, and J. Li, "Leveraging Transfer Learning and GAN Models for OCR from Engineering Documents," in *World AI IoT Congress (AIIoT)*. IEEE, 2022, pp. 015–021.

[36] A. Bhatti, A. Arif, W. Khalid, B. Khan, A. Ali, S. Khalid, and A. u. Rehman, "Recognition and classification of handwritten urdu numerals using deep learning techniques," *Applied Sciences*, vol. 13, no. 3, p. 1624, 2023.

[37] S. B. Ahmed, S. Naz, S. Swati, and M. I. Razzak, "Handwritten urdu character recognition using one-dimensional blstm classifier," *Neural Computing and Applications*, vol. 31, no. 4, pp. 1143–1151, 2019.

[38] A. Mostafa, O. Mohamed, A. Ashraf, A. Elbehery, S. Jamal, A. Salah, and A. S. Ghoneim, "An end-to-end ocr framework for robust arabic-handwriting recognition using a novel transformers-based model and an innovative 270 million-words multi-font corpus of classical arabic with diacritics," *arXiv preprint arXiv:2208.11484*, 2022.

[39] İ. Dölek and A. Kurt, "A deep learning model for ottoman ocr," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 20, p. e6937, 2022.

[40] S. Boudelaa, M. Perea, and M. Carreiras, "Matrices of the frequency and similarity of arabic letters and allographs," *Behavior Research Methods*, vol. 52, pp. 1893–1905, 2020.

[41] Wikipedia, "Arabic letter frequency," https://www.intellaren.com/articles/en/a-study-of-arabic-letter-frequency-analysis, 2023, [Accessed 05-12-2023].

[42] A. El-Sawy, M. Loey, and H. El-Bakry, "Arabic handwritten characters recognition using convolutional neural network," *WSEAS Transactions on Computer Research*, vol. 5, no. 1, pp. 11–19, 2017.

[43] M. Pechwitz, S. S. Maddouri, V. Märgner, N. Ellouze, H. Amiri *et al.*, "IFN/ENIT-database of handwritten Arabic words," in *Proc. of CIFED*, vol. 2. Citeseer, 2002, pp. 127–136.

[44] A. Lawgali, M. Angelova, and A. Bouridane, "HACDB: Handwritten Arabic characters database for automatic character recognition," in *European workshop on visual information processing (EUVIP)*. IEEE, 2013, pp. 255–259.

[45] S. A. Mahmoud, I. Ahmad, W. G. Al-Khatib, M. Alshayeb, M. T. Parvez, V. Märgner, and G. A. Fink, "KHATT: An open Arabic offline handwritten text database," *Pattern Recognition*, vol. 47, no. 3, pp. 1096–1112, 2014.

[46] F. Chabchoub, Y. Kessentini, S. Kanoun, V. Eglin, and F. Lebourgeois, "SmartATID: A mobile captured Arabic Text Images Dataset for multi-purpose recognition tasks," in *International Conference on Frontiers in Handwriting Recognition (ICFHR)*. IEEE, 2016, pp. 120–125.

[47] A. Sulaiman, K. Omar, and M. F. Nasrudin, "A database for degraded Arabic historical manuscripts," in *International Conference on Electrical Engineering and Informatics (ICEEI)*. IEEE, 2017, pp. 1–6.

[48] S. M. Awaidah and S. A. Mahmoud, "A multiple feature/resolution scheme to Arabic (Indian) numerals recognition using hidden Markov models," *Signal Processing*, vol. 89, no. 6, pp. 1176–1184, 2009.

[49] J. Ramdan, K. Omar, M. Faidzul, and A. Mady, "Arabic handwriting data base for text recognition," *Procedia Technology*, vol. 11, pp. 580–584, 2013.

[50] N. E. B. Amara, O. Mazhoud, N. Bouzrara, and N. Ellouze, "ARABASE: A Relational Database for Arabic OCR Systems." *Int. Arab J. Inf. Technol.*, vol. 2, no. 4, pp. 259–266, 2005.

[51] Y. Al-Ohali, M. Cheriet, and C. Suen, "Databases for recognition of handwritten Arabic cheques," *Pattern Recognition*, vol. 36, no. 1, pp. 111–121, 2003.

[52] R. Najam and S. Faizullah, "A scarce dataset for ancient arabic handwritten text recognition," *Data in Brief*, vol. 56, p. 110813, 2024.

[53] R. Najam and Faizullah, "Historical arabic handwritten text recognition dataset, mendeley data," https://data.mendeley.com/datasets/xz6f8bw3w8/1, 2024, [Accessed 16-10-2024].

A Secure Scheme to Counter the Man in the Middle Attacks in SDN Networks-Based Domain Name System

Frank Manuel Vuide Pangop[1], Miguel Landry Foko Sindjoung[2], Mthulisi Velempini[3]
IUT-FV of Bandjoun, University of Dschang, Bandjoun, Cameroon[1]
Department of Computer Science, University of Limpopo, Mankweng, South Africa[2,3]

***Abstract*—Internet and computer networks are vulnerable to cyber-attacks which compromise the services they provide to facilitate the management of data and users. The domain name system (DNS) is the Internet service that translates domain names and computer IP addresses and IPs to domain names. DNS is sometimes a victim of attacks that are difficult to detect and prevent because they are not only very stealthy but also conceal its proper functioning. Among the attacks that DNS is subject to, there are man-in-the-middle (MITM) attacks. Traditional networks that centralize all network functions in a single device complicate the detection and protection of systems against these attacks challenging. Software-defined networking (SDN) is a technology that is widely used to address many traditional network problems such as security and network architectures. Therefore, in this paper, we propose a scheme designed to detect and block man-in-the-middle attacks based on a newly defined architecture. The effectiveness of our secured solution is evaluated in an SDN architecture where an Address Resolution Protocol spoofing MITM attack is generated for the evaluation purpose. The results of our simulations show that we can effectively detect the attack and the performance evaluation of our approach shows that the proposed solution is effective in terms of security, implementation cost and resource consumption. We then recommend the use of our proposed solution to address the MITM attacks in SDN networks-based Domain Name System.**

***Keywords*—Cyber security; domain name system; man in the middle attack; software defined networking**

I. Introduction

Cybersecurity protects computer systems from malicious attacks. Cybersecurity [1] is the set of measures and practices which protect computer systems, networks, data and users against attacks, intrusions and online threats. Interestingly, human error counts for 50-80 % of network outages [2] due to misconfigurations made by administrators when configuring a network. These attacks also affect the domain name system (DNS).

The DNS is a hierarchical, distributed system for associating domain names with IP addresses. DNS is also vulnerable to attacks [3]. New forms of attacks that are stealthy and sophisticated [4] are difficult to address in traditional networks. The use of new technologies, such as software-defined networking (SDN) is a new networking paradigm that facilitates network management and administration by providing an interface to control network infrastructure such as switches which can be employed to address the security problem.

DNS servers are victims of several attacks because of their role and visibility on the Internet. Some of the common attacks are amplification and DoS attacks, botnets and attacks using DNS, DNS Zone transfer attacks, and DNS manipulation. DNS manipulation compromises DNS resolvers responsible for manipulating DNS responses [5]. In recent years, several researchers have focused on finding ways to counter DNS manipulation that uses active measurements to profile open resolvers. In this paper, we focus on DNS manipulation, more precisely on the man-in-the-middle (MITM) attack.

The MITM attack is an interference where an attacker eavesdrops, intercepts, or manipulates communication between two or more parties to steal information. In DNS, the attacker manipulates DNS servers or cache to redirect users to fraudulent sites to steal information or download the malware [6]. The attacker using a MITM attack intercepts communication on the communication channel and pretends to be a receiver to the sender and the sender to a receiver. This can be possible following two scenarios: firstly, communication between the two entities is not secure, in other words, the start of the communication is not preceded by an authentication; secondly, the two parties communicate over an insecure communication channel, such as public Wi-Fi or an unencrypted network. So, an attacker can intercept and manipulate the messages.

It is, therefore, necessary to find effective mechanisms to prevent and detect MITM attacks in DNS to ensure the security and confidentiality of users' communications. This involves implementing protection mechanisms such as encryption of communications, mutual authentication, Intrusion Detection System/Intrusion Prevention System (IDS/IPS), and the use of protocols such as Secure Socket Layer/Transport Layer Security (SSL/TLS). In this article, we propose a solution to detect and block MITM attacks on DNS servers based on a new SDN architecture.

The rest of this paper is organized as follows: in Section II, we present the related work. We then present our proposed scheme to address the MITM attacks in SDN networks-based DNS in Section III. Thereafter, we present the simulation results and analysis in Section IV. Finally, we conclude the paper in Section V.

This work is based on the research supported by the National Research Foundation of South Africa (Grant Numbers: 141918).

II. State of the Art

In the literature, many solutions have been proposed to counter security attacks in SDN architecture. In this section, we review some of the related works.

In study [7], Sungmin et al., proposed a new security extension named Topoguard to the existing OpenFlow controllers to provide automatic and real-time detection of network topology poisoning attacks in an SDN architecture. Their solution is premised on the poisoned network visibility and the upper-layer OpenFlow controller, services/applications may be misled, leading to hijacking, denial of service or MITM attacks. Topoguard is based on SDN/OpenFlow topology service security analysis for vulnerability identification. Sungmin et al. have proposed a network topology poisoning attack to exploit the vulnerabilities their solution has identified. Moreover, they investigated the defence space and proposed an automatic mitigation approach against network poisoning attacks and a prototype defence system.

In study [8], Cheng et al. investigated the potential threats that the OpenFlow control channel may face from the MITM attacks. They proposed an attack model in an IoT-Fog architecture and they also demonstrated the consequences of these attacks. Based on the previous demonstration, Cheng et al. [8] proposed a lightweight countermeasure based on Bloom filters. Anass et al. [9] proposed a Context-Based Node Acceptance (CBNA) framework to mitigate the MITM threats based on the authentication of new nodes using OpenFlow switches in a software-defined network. The CNBA framework takes place when new nodes attempt to connect to the controller for the first time and make a decision based on the response time of nodes.

Sahri et al. [10] proposed a collaborative SDN authentication (CAuth) which blocks the spoofed packet while authenticating the legitimate packet when establishing communications between a client and the DNS server within an SDN architecture. Their solution allows authentication of the two entities wishing to communicate with each other (client and DNS). The authentication is made between the client controller and the DNS server controller. When the DNS server controller receives the query packet, it sends an authentication packet to the clients that initiate the communication; when the DNS server controller receives an authentication from the client, it replies with the DNS reply to the client. During this exchange, the switch flow table is updated from a match check with the incoming packet to detect the attacking packets. This method can discriminate legitimate packets from attacking ones. Moreover. it can block the attacks before they reach the DNS server, but it consumes a lot of bandwidth. Main focus in this paper will be on the MITM attacks

III. The Proposed Scheme to Address the MITM Attacks in DNS based on SDNs

In this section, we present our contribution to counter the MITM attacks in DNS based on SDN networks. We present a SDN architecture (Section III-A) within which our proposed secure MITM attacks algorithm (Section III-B) is executed.

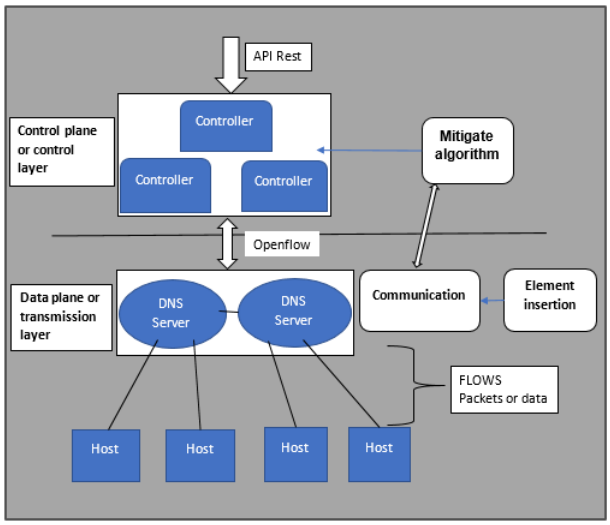


Fig. 1. The proposed architecture.

A. The Proposed Architecture

The proposed architecture is organised into two main layers: the infrastructure layer and the control layer. These two layers are linked by a communication interface which promotes interaction between them. The proposed architecture is shown in Fig. 1 and includes some modules at each layer. The communication interface acts as a bridge between the infrastructure and the control layers. Its main role is to ensure data transfer from one layer to another. The roles of each layer are discussed below:

1) The infrastructure layer: The infrastructure layer is responsible for collecting and transmitting data to the post traffic layer in the network for analysis. The data collection is done according to the rules defined by the control layer. Importantly, this layer has the various network equipment that are DNS server and hosts as shown in Fig. 1. The following modules are incorporated in the infrastructure layer:

- The communication module: It is used to establish communication rules between the DNS server of the infrastructure layer and the controller in the control layer through the communication interface. The collected data is transmitted to the control layer, otherwise, no verification will be made at that layer for attack detection.
- The element insertion module: It is responsible for filtering the data stream collected from the parameters defined by the controller. If these parameters are not defined, then all the network data flow is sent to the control layer, resulting in the increase of the workload in the controller and consequently, the analysis of some incoming data may be ignored, causing a security risk in the network.

2) The control layer: This layer is responsible for overall network monitoring. In other words, it allows the management and control of the flow of data transmitted in the network. This layer works through the controller. The controller fulfils or applies all the functions of the control layer and serves as a

bridge between the applications and the infrastructure layer. Our proposed algorithm (Algorithm 1) which mitigates the MITM attacks is executed in this control layer.

B. The Proposed Algorithm

To detect the MITM attacks in DNS based on SDN network, we propose Algorithm 1 named MITM attack detection (MitAL). MitAL works as follows: Firstly, the rules for DNS-controller communication are established. After the traffic is generated, the transmitted packets are captured and sent to the controller for analysis. If the packet is of ARP type then an attack is detected. If the MAC address of the packet is not the expected one or there is no corresponding MAC address in the MAC-IP table, it is classified as malicious. The MAC address received with the incoming packet is considered as MAC address of a malicious node. Therefore, a notification is sent to all the DNS servers of the network to block all communications from the host with the detected malicious MAC address. Moreover, if the packet is of IP type then an attack is detected if the IP address is valid and the three following conditions are satisfied:

1) The IP address of the node is external to the logical network
2) The response time is greater than the defined threshold value, and
3) The received IP address is not the expected one.

In the previous scenario, a notification is sent to all the DNS servers of the network to block all the communications from the host with the detected malicious IP address.

Algorithm 1: MitAL:MITM attack detection

```
Input: Thr: Threshold in milliseconds, ip_to_mac = ():
       MAC-IP address table;
Capture packets after generating network traffic;
Retrieve information (types) about each incoming
  packet;
For each packet P Do
    If ARP packet Then
        Extract (src_mac, src_ip) of packet;
        If ip_to_mac[P.src_ip] != P.src_mac Then
            Block all the communications from the host
              with MAC address src_mac
    Else
        If IP packet Then
            Extract (src_ip, dst_ip) of packet;
            If src_ip is valid Then
                time ← Calculate response_time();
                mask ← Check if is_external (src_ip) to
                  network;
                source ← Check if src_ip !=
                  expected(src_ip);
                If (time ≥ Thr) and mask and source Then
                    Block all the communications from
                      the host with ip address src_ip
```

To evaluate the effectiveness of our scheme, we perform some simulations where we compare our algorithm to those presented by authors in studies [9] and [10]. The generated results are presented in Section IV.

IV. Simulations and Analysis

This section presents the results of our simulations. These simulations were performed using Mininet* as a network emulator because it creates a network of virtual hosts, switches, controllers and links, and therefore it can be used to build SDN. As a network security tool for analysing, penetration testing and protocol analysis, we have used Ettercap†. Ettercap is a comprehensive suite of tools that allows users to sniff, intercept, and manipulate network traffic in real time. It was used in our simulations to generate the MITM attacks. Wireshark‡ was also used to capture and analyse network traffic in real time. It can support a range of protocols and various capture interfaces. Table I summarizes the details of the used simulation environment and Fig. 2 shows testbed environment.

TABLE I. Simulation Environment

Tool	Version
OS	Ubuntu 22.04 LTS
Python	3.10
Mininet	2.3.0
Ettercap	0.8.3.1
Wireshark	3.6.2
Ryu controller	4.34

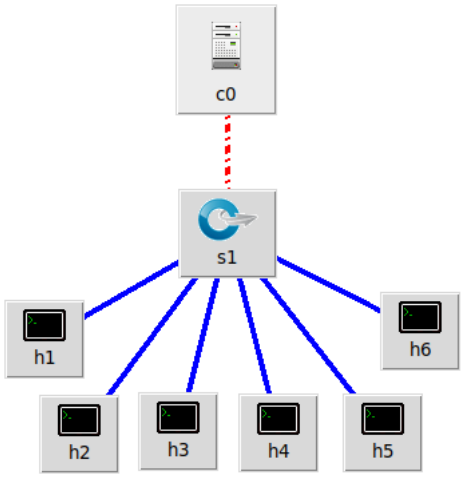


Fig. 2. Testbed environment.

We have used the Mininet generator to create SDN network topology on Ubuntu 22.04 LTS installed on a laptop with an intel core i3 processor, RAM - 4GB, Hard disk - 30GB. We used the Ryu controller for network management and monitoring (We justify the choice of the controller in Section IV-A). The SDN network set-up includes a Ryu controller, one open vSwitches (s1) acting as DNS server, and six end-hosts (h1-h6). The non-malicious hosts are h3 and h4, and the malicious host is h1. To generate normal traffic between non-malicious hosts, we used iperf3 and ping command and we generated an attack (ARP spoofing - MITM attack) traffic using Ettercap. The use of Wireshark assisted us in retrieving more details or information about the attack.

*http://mininet.org
†https://www.ettercap-project.org
‡https://www.wireshark.org

A. Choosing the Suitable Controller for our Simulations

To choose the suitable controller to be used in our simulations, We evaluated three existing controllers designed for mininet environments: Ryu, Pox and Floodlight. The evaluation was inspired by the work presented in study [11]. The parameters used for comparison are presented in Table II.

The evaluation was done by generating traffic and varying the number of switches from 5 to 100 in order to obtain accurate results. During this time, data related to bandwidth and latency were collected and the resulting graphs are presented in Fig. 3(a) and Fig. 3(b). The evaluated controllers are Ryu, Pox and Floodlight. In Fig. 3(a), we observed that overall, the Pox controller consumes more bandwidth compared to the Ryu and Floodlight controllers, especially when the number of switches is less than 64. At the same time, the Ryu controller consumes more bandwidth than Floodlight. However, in Fig. 3(b), we observed that the latency is overall the best when using the Ryu controller compared to the Pox (medium) and Floodlight (Bad) controllers. In conclusion, among the evaluated controllers, Ryu is moderate in bandwidth consumption and the best for latency. For this reason, it was selected and used in our simulations.

B. Attack Detection using Flow Table

After traffic is generated by an attacker, the protocol stores any information about each incoming packet according to different fields defined in the flux table. By inspecting this table and the contained information, we were able to detect the presence of a MITM attacker in the communication between the two non-malicious hosts as illustrated in Fig. 4. It shows the contents of the flow table before and after attack generation. Before the attack, normal traffic between the non-malicious hosts is observed without interference in their communications. After the attack is launched, the attacker initiates, intercepts and manipulates the communication between the legitimate hosts by impersonating one of the two entities according to the sent packet. The highlighted green and red areas respectively represent the non-malicious and the attacking hosts.

C. Capturing Network Traffic using Wireshark

We used wireshark for overall network capture and analysis to detect the attack and to have more details and information about the attack. Fig. 5 illustrates a wireshark capture of all traffic generated in the network. We can observe normal communication between network hosts using the ICMP protocol. An example of attack traffic is observed by the area highlighted in red where we note a host which initiates communication between the two non-malicious nodes to intercept and manipulate the traffic. We also observe that the protocol used is modified in ARP since in the ARP spoofing attack, the size and the information on the packets are modified.

Fig 6(a) and 6(b) illustrate the content of a request and a response sent from the attacking host; when communicating with a non-malicious host. They show the request and reply content from the attacker. Each of these messages contains a header and a body specific to it. The header allows to see some information about the message in transit. That is, the detection of an attacking host is effective, both in packet transmission and reception.

D. Performance Based on Bandwidth and Latency

Bandwidth is used to determine a network's ability to transfer information/packets over a defined period. Generally expressed in Mbits/sec (can also be expressed in Kbits/s or Gbits/sec), in our work, we expressed it in Gbits/sec. Latency in a network is the time it takes for an entire data packet to be sent from one point to another. It is measured in milliseconds (ms) and can be influenced by various factors such as the distance between the two points, the quality of the connection, and the number of hops between the points. To compare our algorithm according to bandwidth and latency in the network, we simulated an exchange of packets and collected the associated data to represent it in graphs illustrated in Fig. 7(a) and 7(b).

Fig. 7(a) depicts the comparative bandwidth results of our algorithm (MitAL) and the work of Anass et al.[9] (CBNA). The results show that: for five (5) switches the CBNA consumes less bandwidth than MitAL, however, for $n = [10, 20, 32, 64, 100]$ where n is the number of switches, MitAL consumes less bandwidth than CBNA. From this comparison, we can conclude that our algorithm minimises the bandwidth consumption.

Fig. 7(b) shows the latency between our algorithm (MitAL) and the work of CBNA framework of Anass Sebbar et al. [9]. The graph obtained from the collected data illustrates that for $n = [5, 10, 32, 100]$ where n is the number of switches, the latency occurred by MitAL is better than that of CBNA. This comparison show that our algorithm minimises the latency since in most of the considered points our algorithm achieved better performances.

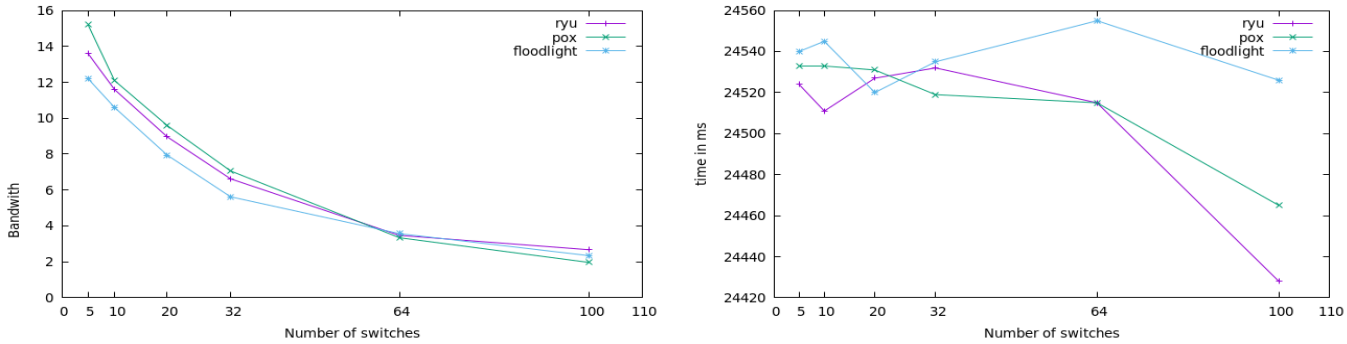
E. Traffic without Protection vs. Traffic with MitAL

Fig. 8 shows the comparison between the network traffic without protection and the traffic with our proposed solution (MitAL). We can observe and conclude that when the network is not protected and the attack is generated, the packet transmission time is high which means that the latency is high. Unlike with our algorithm and the associated security protocols, we observe that this time is considerably reduced. This shows the effectiveness of our protocol.

F. Detection Rate by Number of Generated Attacks

To evaluate the performance of our algorithm, we used the attack detection rate parameter. This rate is determined from the number of generated attacks, and it shows the performance in terms of detection when the MITM attack generated from different sources. It is computed using evaluation metrics such as throughput, available bandwidth when the attack is generated, and analysis of network communications to determine healthy packets from the number of packets exchanged, attacked and lost (percentage).

We have compared MitAL to the solution by Anass et al. (CBNA) [9] and that of Sahri et al. (CAuth) [10]. The comparison graph is provided in Fig. 9. When analysing that figure, we can conclude that our solution outperforms CBNA and CAuth when the number of generated attacks increases until it reaches 60. From that point, the detection rate is still better for MitAL compared to CAuth, but less than that of CBNA, which is quite satisfactory.



(a) A comparative test based on bandwidth

(b) A comparative test based on latency

Fig. 3. A comparative study between controllers that may be considered.

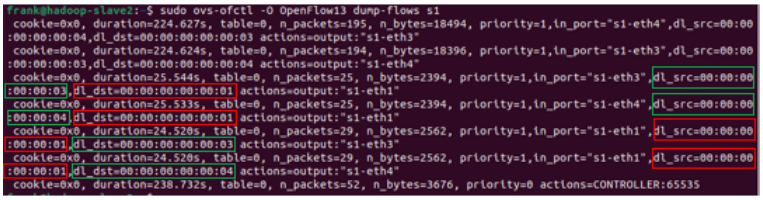


Fig. 4. Communication handling by attacker.

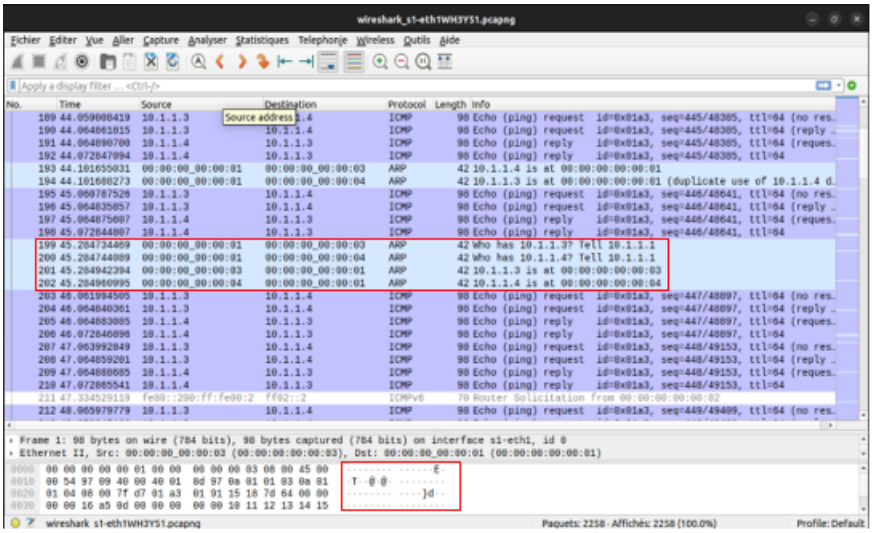
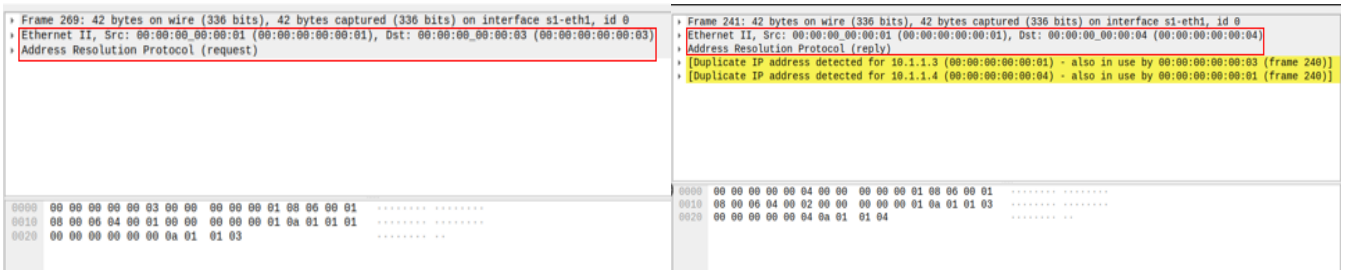


Fig. 5. Packets capture in wireshark.



(a) Capturing request body from attacker

(b) Capturing reply body from attacker

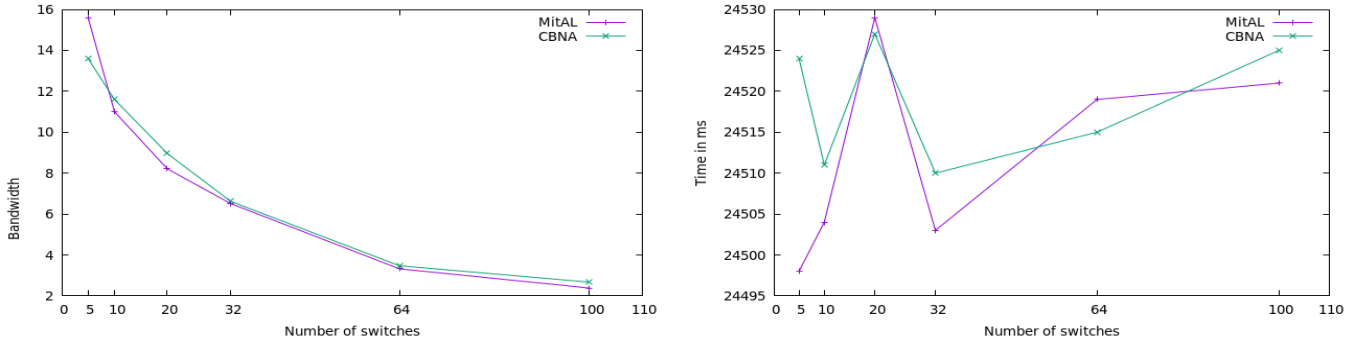
Fig. 6. Data captured from attacker.

TABLE II. PARAMETERS USED FOR TESTING

Parameters	Bandwidth	Latency
Topology type	Linear (n switches, n hosts)	Linear (n switches, n hosts)
IP address range for hosts	10.0.0.1-10.0.0.n	10.0.0.1-10.0.0.n
Traffic generating tool	iperf3	ping
Graph generating tool	gnuplot	gnuplot

TABLE III. COMPARATIVE STUDY BETWEEN SOME EXISTING PROTOCOLS

Protocol	Type of attacks	Security mesure	Security protocols	Detection rate
Sahri et al. [10] (CAuth)	IP Spoofing	Authentication	TLS	75%
Anass et al. [9] (CBNA)	MITM	Latency	TLS	82%
Our protocol (MitAL)	MITM	Authentication, Latency, Cryptographic Keys	TLS, SSL	86%



(a) Evaluation based on bandwidth

(b) Evaluation based on latency

Fig. 7. Performance evaluation of our algorithm.

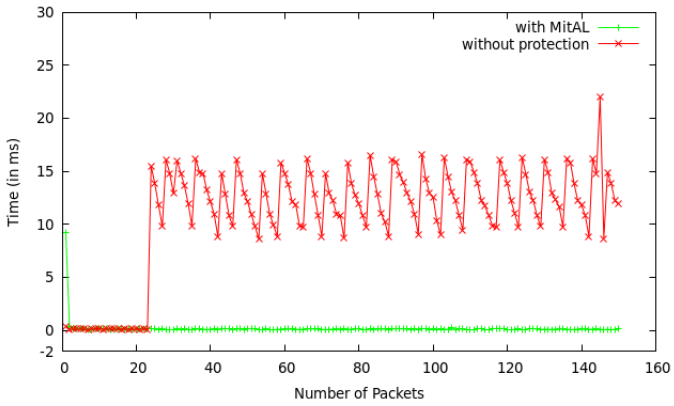


Fig. 8. Traffic comparison.

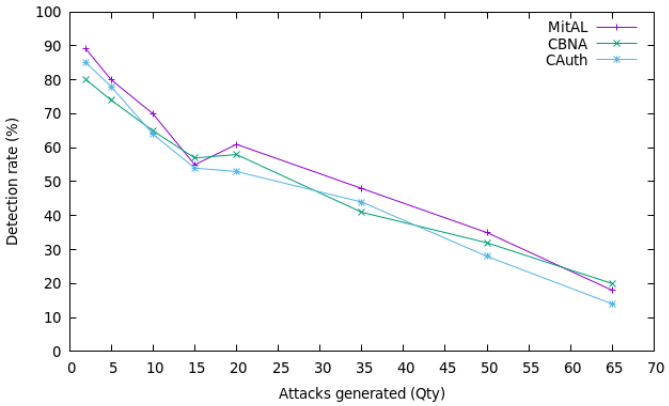


Fig. 9. Detection rate by number of attacks generated.

```
Detected attacker. Adresse IP of node :  10.1.1.1
Node with MAC address 00:00:00:00:00:01 and IP address 10.1.1.1 is an attacker
Attack detected and Prevent to port : 1
```

Fig. 10. System response to block attacker.

G. Discussion

In the previous sections, we presented results that we generated through simulations. The results show that the detection of malicious nodes is efficient. Secondly, the results provide detailed information about the attacks (the MAC addresses, the protocol used, and the data contained in the packets). Once an attack is detected, the protocol displays information about the attacking node and builds a flow rule to block its communication in the network by blocking its connection port as shown in Fig. 10.

As presented in Table III, our proposed MitAL protocol for DNS servers in an SDN architecture aims to address MITM attacks like CBNA [9] unlike CAuth [10] that address the IP Spoofing attacks. The secure methods used by MitAl are authentication, latency and cryptographics keys while CBNA

only uses latency and CAuth uses authentication. Furthermore, CAuth and CBNA are based on TLS as a security protocol while MitAl uses both TLS and SSL. Finally, it should be noted that MitAL has a detection rate of 86%, which is higher than the 82% of CBNA and the 75% of CAuth protocols.

V. Conclusion

The DNS is the victim of many kinds of attacks that are difficult to detect and prevent. Among these attacks, are MITM attacks. MITM attacks cause several problems in real-life networks such as tampering, integrity violation and exposure or divulging of confidential or user data. In this paper, we proposed a scheme which is based on SDN architecture. The scheme detects and blocks the MITM attacks on DNS servers. We first presented our designed SDN architecture then we proposed an algorithm that is implemented by the SDN controller to guarantee the security of communications between different trusted hosts of the network, and finally, we presented the simulations results we generated through simulations. The results show that our secure algorithm is effective and efficient in detecting MITM attacks. A comparison with other existing algorithms shows that our proposed solution is superior. We therefore recommend our proposed algorithm (MitAL) as a suitable candidate for use to mitigate the effects of MITM attacks on the DNS server. Despite our results being satisfactory, we believe that the use of Machine Learning techniques can significantly improve the efficiency of our solution. This aspect will be explored in detail in future

References

[1] N. R. Gade and U. Reddy, "A study of cyber security challenges and its emerging trends on latest technologies," 02 2014.

[2] S. Matsumoto, S. Hitz, and A. Perrig, "Fleet: defending sdns from malicious administrators," 08 2014.

[3] J. C. C. Chica, J. C. Imbachi, and J. F. B. Vega, "Security in SDN: A comprehensive survey," *Journal of Network and Computer Applications*, vol. 159, p. 102595, 2020.

[4] J. Zheng, Q. Li, G. Gu, J. Cao, D. K. Y. Yau, and J. Wu, "Realtime ddos defense using cots sdn switches via adaptive correlation analysis," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 7, pp. 1838–1853, 2018.

[5] M. Kuhrer, T. Hupperich, J. Bushart, C. Rossow, and T. Holz, "Going wild: Large-scale classification of open dns resolvers," in *Proceedings of the 2015 Internet Measurement Conference*, ser. IMC'15. New York, NY, USA: Association for Computing Machinery, 2015, pp. 355–368. [Online]. Available: https://doi.org/10.1145/2815675.2815683

[6] A. Khormali, J. Park, H. Alasmary, A. Anwar, M. Saad, and D. Mohaisen, "Domain name system security and privacy: A contemporary survey," *Computer Networks*, vol. 185, p. 107699, 2021.

[7] S. Hong, L. Xu, H. Wang, and G. Gu, "Poisoning network visibility in software-defined networks: New attacks and countermeasures," in *Ndss*, vol. 15, 01 2015, pp. 8–11.

[8] C. Li, Z. Qin, E. Novak, and Q. Li, "Securing sdn infrastructure of iot fog networks from mitm attacks," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1156–1164, 2017.

[9] A. Sebbar, K. Zkik, M. Boulmalf, and M. D. E.-C. El Kettani, "New context-based node acceptance cbna framework for mitm detection in sdn architecture," *Procedia Computer Science*, vol. 160, pp. 825–830, 2019.

[10] N. Sahri and K. Okamura, "Protecting dns services from ip spoofing: Sdn collaborative authentication approach," in *Proceedings of the 11th International Conference on Future Internet Technologies*, ser. CFI '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 83–89.

[11] O. Salman, I. H. Elhajj, A. Kayssi, and A. Chehab, "Sdn controllers: A comparative study," in *2016 18th mediterranean electrotechnical conference (MELECON)*. IEEE, 2016, pp. 1–6.

Efficient Load-Balancing and Container Deployment for Enhancing Latency in an Edge Computing-Based IoT Network Using Kubernetes for Orchestration

Garrik Brel Jagho Mdemaya[1], Milliam Maxime Zekeng Ndadji[2],
Miguel Landry Foko Sindjoung[3], Mthulisi Velempini[4]
Department of Computer Science, University of Limpopo, Mankweng, South Africa[1,3,4]
Department of Mathematics and Computer Science, University of Dschang, Dschang, Cameroon[2]

***Abstract*—Edge Computing (EC) provides computational and storage resources close to data-generating devices, and reduces end-to-end latency for communications between end-devices and the remote servers. In smart cities (SC) for example, thousands of applications are running on edge servers, and it becomes crucial to manage resource allocation and load balancing to improve data transmission throughput and reduce latency. Kubernetes (k8s) is a widely used container orchestration platform that is commonly employed for the efficient management of containerized applications in SC. However, it does not integrate well with certain EC requirements such as network-related metrics and the heterogeneity of EC clusters. Furthermore, requests are equally distributed across all replicas of an application, which may increase the time taken for processing, since in the EC environment, nodes are geographically dispersed. Several existing studies have investigated this problem, unfortunately, the proposed solutions consume a lot of node's resources in the cluster. To the best of our knowledge, none of studies considered the cluster heterogeneity when deploying applications that have different resource requirements. To address this issue, this paper proposes a new technique to deploy applications on edge servers by extending Kubernetes scheduler, and an approach to manage requests among the different nodes. The simulation results show that our solution generates better results than some of the state-of-the-art works in terms of latency.**

***Keywords*—Latency; Kubernetes; edge computing; Internet of Things; load-balancing**

I. Introduction

The fusion of Internet of Things (IoT) and edge computing has revolutionized distributed computing, enabling real-time data processing and analysis at the network edge in metropolitan cities [1, 11]. Various technologies are enhancing urban life. Numerous fundamental departments of cities such as transportation, power plants, information systems, crime detection traffic monitoring, etc. are equipped with sensors to gather data on community services. The endeavour to construct a smart city entails an increasing number of sensors and a growing volume of traffic. This evolution also impacts applications operating within smart city-related frameworks. Thus, thousands of applications are running in real-time on nodes at the edge layer. So they need to be organised well and orchestrated to serve efficiently each activity sector. Kubernetes (k8s), a leading container orchestration platform, has extended its reach to the edge, offering sophisticated tools for managing containerized applications in this decentralized environment [15, 3]. However, orchestrating containers on heterogeneous edge clusters poses significant challenges. Edge nodes come in diverse forms, ranging from resource-constrained sensors to powerful edge servers, each with distinct processing capabilities and network connectivity. Also, the edge layer is made up of several heterogeneous machines forming a cluster. Some machines are much more powerful than others; consequently, certain processes, such as artificial intelligence training, need to be run on much more powerful machines, while certain processes, such as saving and collecting multimedia data, can be run on less powerful machines.

Despite heterogeneity, smart deployment of applications and efficient utilization of computing resources require the distribution of incoming requests across machines within the cluster. These tasks fall under the kube-scheduler which is the component in charge of application deployment on nodes in a Kubernetes cluster, and under the ingress controller and kube-proxy components, which play a significant role in balancing the workload and optimizing performance across heterogeneous edge nodes. Several existing works designed to improve latency and throughput in edge computing-based IoT networks using Kubernetes as a container orchestration tool have been proposed in the literature. However, some of the proposed techniques designed to extend the default kube-scheduler consume a fair amount of energy due to the workload on nodes [12, 9]. Others use custom orchestrators running in containers in the same cluster [20, 7], which consumes the hardware resources of machines whose main aim is to respond as quickly as possible to end-user requests. Some use strategies to distribute requests fairly among the nodes in the cluster using the kube-proxy component. This does not guarantee a significant reduction in latency given that they do not consider some node hardware resources during the load-balancing process [14]. In this paper, to address the latency problem in an Edge computing-based IoT network using Kubernetes for container orchestration, we first introduce a solution to deploy applications on edge nodes depending on the resources required by these applications to run smoothly on the cluster nodes. Secondly, we propose an approach to manage load-balancing and distribute the requests among the nodes of the cluster while considering hardware resources available on cluster nodes. Thus, the contribution of this paper is summarized as follows

This work is based on the research supported by the National Research Foundation of South Africa (Grant Numbers: 141918)

- We show how the default kube-scheduler can be extended to evaluate and provide a score to applications and deploy them on edge nodes in an efficient way while considering the type of nodes and the available resources.
- We explore the complexities of container orchestration at the edge using Kubernetes, emphasizing the importance of accommodating cluster heterogeneity and implementing request distribution mechanisms for better load-balancing.

The remainder of this paper is organized as follows: Section II reviews the related works. Section III presents our approach to addressing the latency problem, while in Section IV we present the simulation results and discuss the performance of the proposed approach. Finally, Section V concludes the paper and discusses the perspectives of our proposal.

II. Literature Review

This section aims to present the works that are related to our study. We start by presenting kubernetes in Section II-A, then, the state of the art on works that use Kubernetes as orchestrators to reduce the latency in Section II-B.

A. What is on Kubernetes?

Kubernetes is a prominent open-source platform designed to streamline the deployment, management, and scaling of containerized applications. A Kubernetes cluster comprises of master nodes (or control planes) and worker nodes. Applications are run in units called pods, that serve as Kubernetes, the smallest execution units [16]. The control plane manages the worker nodes and the pods within the cluster. In production environments, the control plane operates across multiple computers, and a cluster generally encompasses multiple nodes, ensuring fault tolerance and high availability. Fig. 1 shows the kubernetes cluster architecture.

The control plane components are responsible for making overarching decisions regarding the cluster, such as scheduling, identifying and reacting to cluster events, such as initiating a new pod when necessary. These components have the flexibility to operate on any machine within the cluster. However, to streamline operations, setup scripts commonly initiate all control plane components on a single machine, and also avoid executing user containers on a machine. The control pane has four main components: (1) **kube-apiserver** is a key component also known as API server for the Kubernetes control plane that serves as an interface for the Kubernetes API and acts as its primary gateway. (2) **etcd** is a database within which Kubernetes information like metadata, current state and desired states are stored. (3) **kube-scheduler** is the control plane component that monitors created pods without assigned nodes and selects suitable nodes for their execution. Finally, (4) **kube-controller-manager** is the control plane component responsible for executing controller processes.

The worker nodes are responsible for the applications (pods) that are running in the cluster. They mainly have three components that run on every worker node, with the main role of maintaining the running pods and providing the Kubernetes runtime environment. (1) **kubelet** is a node-level agent that helps to ensure the operational status of containers within pods. (2) **kube-proxy** operates as a network proxy and serves as a fundamental component of the Kubernetes service framework. Finally, (3) **Container runtime** is an essential element that enables Kubernetes to efficiently run containers.

In a Kubernetes cluster, relying on the IP address of an application pod for access can be challenging due to the pod's IP changing upon restart. To ensure application reachability, a layer of abstraction known as a service is employed to expose groups of application pods to clients. Each service is assigned a virtual IP address, called ClusterIP, which remains unchanged unless recreated, thus guaranteeing application reachability within the cluster. However, ClusterIP is only accessible within the cluster. To enable access from outside the cluster, services can be configured with NodePort or LoadBalancer. With NodePort, a static port is opened on each node, allowing access through nodeIP:port from inside or outside the cluster. LoadBalancer leverages the cloud provider's load-balancing mechanisms to externally expose the service. Traffic accessing an application is routed to its pods and monitored by the service, with routing decisions based on kube-proxy modes such as userspace, iptables, and IPVS.

By default, kube-proxy operates in userspace mode, utilizing a round-robin algorithm for traffic distribution. Alternatively, iptables mode randomly selects a pod for traffic handling. For large-scale applications, IPVS offers more efficient routing algorithms including round-robin, least connection, destination hashing, source hashing, shortest expected delay, and never enqueued. Another important component in Kubernetes is the horizontal pod autoscaler (HPA), which is in charge of creating automatically new pods of an application if the existing pods are overloaded, or deleting some of them if the application is not being used.

B. Overview of Works using Kubernetes as Orchestrator

Many recent studies have focused on the integration of Kubernetes for load-balancing management in edge computing-based IoT networks. Although Kubernetes has some limitations in edge computing-based IoT networks, the works presented in study [4, 19] show that it remains a promising candidate for managing load balancing and improving latency and throughput. Cilic et al. [5] found that Kubernetes and its distributions hold promise in facilitating efficient scheduling across edge network resources. Nevertheless, several challenges remain to be addressed to optimize these tools for the dynamic and distributed execution environment inherent in edge computing. In research [10], Han et al. introduced Kais, a learning-based scheduling framework for such edge-cloud systems to improve the long-term throughput rate of request processing. Unfortunately, in their work, some nodes have a high workload, while others are under-utilised. Goethals et al. [8] proposed FLEDGE, A container orchestrator compatible with kubernetes, leveraging Virtual Kubelets, primarily targeting container orchestration on resource-constrained edge devices. They did not integrate the fact that edge nodes are most of the time heterogeneous, and in their implementation, they also did not consider the distribution of user requests among the nodes, which does not allow them to achieve better results in terms of latency and bandwidth. Phuc et al. [16] proposed an extension of the default HPA whose role is to offer the

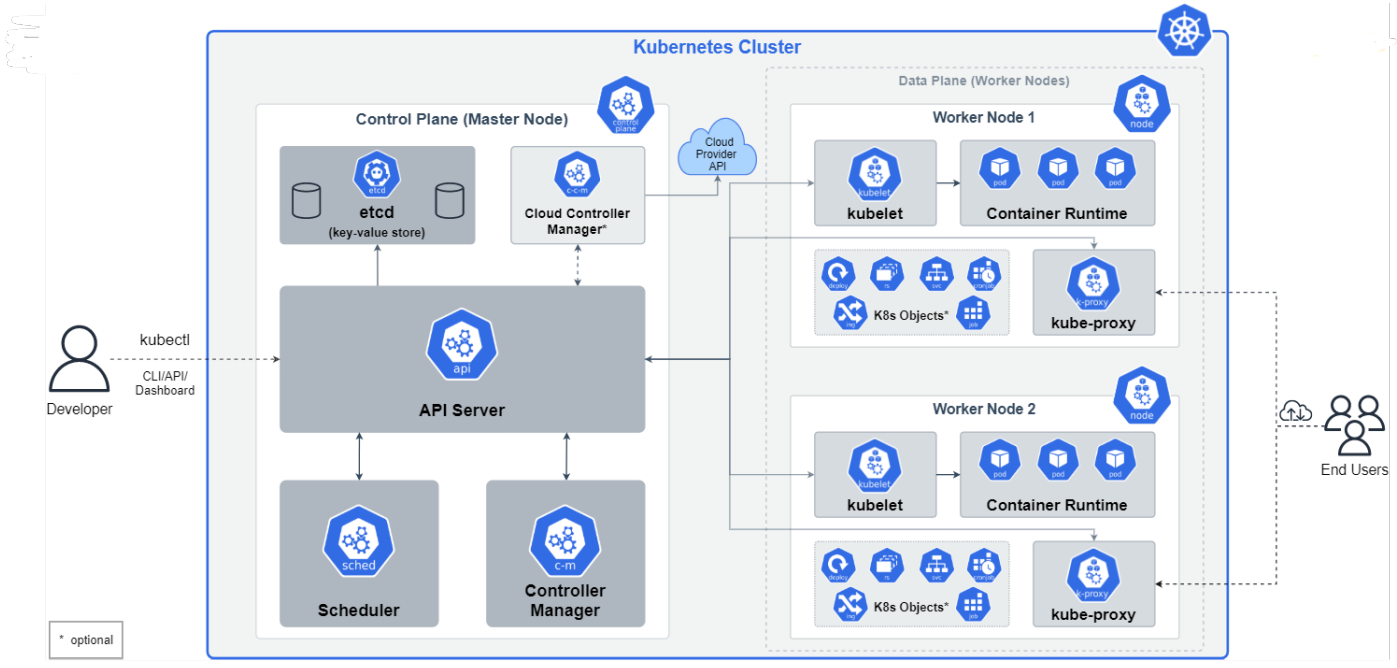


Fig. 1. Kubernetes cluster architecture [2].

resource autoscaling capability. However, the HPA allocates pods to worker nodes without considering the resource demand imbalances in edge computing environments. Authors in [16] considered upscaling and downscaling actions based on network traffic information from nodes to enhance the load-balancing of IoT services in the edge computing infrastructure. Their solution increased throughput and response time between edge nodes, but it doesn't guarantee that a pod that required a lot of resources will be executed on a worker node having the requirements to run that kind of application; therefore, if a particular pod is running on the wrong node, it can increase energy consumption and end-users requests can take too much time to be executed, resulting in high latency. In study [13], the authors studied KubeEdge, which is an open-source platform based on Kubernetes, that aims to orchestrate containerized IoT application services in IoT edge computing environments. In study [6], authors proposed a multi-application hierarchical autoscaling framework tailored for Kubernetes edge clusters. A mechanism based on applications nominates optimal deployment choices through workload prediction and various criteria to ensure application performance while minimizing infrastructure provider costs. This strategy achieves significant improvement in the average allocated resources and energy consumption but it doesn't operate during the first deployment of the application on the servers. Nguyen et al. in study [14], proposed a scheme named Resource Adaptive Proxy (RAP) that considers metrics like latency between worker nodes, Central Processing Unit (CPU) and Random Access Memory (RAM) on worker nodes in order to efficiently distribute the end users requests among the worker nodes. Unfortunately, RAP algorithm does not monitor other cluster resources such as graphics processing units and storage. Rac et al. in STUDY [18] introduce a methodology aimed at reducing the operational costs of applications across the edge-to-cloud computing continuum. Authors in study [17] proposed EdgeOptimizer, a solution that serves as an online testbed for verifying kubernetes-based algorithms, offering detailed configuration options to facilitate cluster management. Unfortunately, their solution does not evaluate the hardware resources consumed by the applications running on their most complex use cases in order to efficiently schedule their deployments in Kubernetes clusters in an edge computing environment.

So far, several works that aim to integrate Kubernetes distributions for container orchestration in edge computing-based IoT networks have been presented. The purpose of doing that is to reduce the latency by managing the load-balancing in edge computing-based IoT networks when considering all the requirements of the edge computing environment. Unfortunately, despite the effectiveness of these solutions, there are still some limitations that need to be addressed while considering Kubernetes for container orchestration in edge computing-based IoT networks.

Edge nodes are heterogeneous and some of them can run faster, some resource-constrained applications than others, and it is therefore important to schedule application deployments by considering the best nodes on which they will efficiently run. Secondly, since Kubernetes and recent works do not integrate network-related metrics and resources like Graphical Processing Unit (GPU) and storage for load-balancing properly. It is important to enhance them and propose solutions to achieve better latency and throughput. Finally, default Ingress-Controller and Kube-proxy components distribute evenly end-users requests among the pods without considering the available hardware resources on the worker nodes. In this work, we address these issues and propose a solution that schedules deployment of containerized applications on worker nodes efficiently, and secondly manages load-balancing by distributing requests among the worker nodes while considering hardware metrics and effective locations of the pods on the worker nodes.

III. Contribution

This section presents our contribution for enhancing latency between end-users and edge layer in edge computing-based IoT networks deployed in smart cities using Kubernetes for container orchestration. The strategy we are proposing consists, on the one hand, extension of the default kube-scheduler installed on the master, to enable it to deploy pods while taking into account the minimum hardware resources required to run those pods and the types of machines in the cluster having these hardware resources. On the other hand, enhancing the default Kubernetes load-balancing algorithms and integrating other metrics with CPU and RAM to manage the workload on the edge layer nodes.

A. Assumptions and Notation

The following assumptions are done for this work:

- There are several edge nodes geographically dispersed in a smart city, and thousands of applications are running on them;
- All applications are containerised using docker, and Kubernetes is used for container orchestration;
- The master node has powerful resources and can launch virtual machines for a sandbox environment, which is a dynamic area where developers can play with codes without fear of disrupting the larger system and where they can test their solutions;
- The worker nodes in the cluster are heterogeneous;
- It is possible to evaluate the hardware resources consumed by a running container and assign a score to that container depending on the resources it consumes;
- It is also possible to evaluate a machine and assign it a score according to its resources;

In Table I, we have a list of notations that will be used to describe our solution

B. Kube-scheduler Awareness

We propose an extension to the default kube-scheduler installed on the master node. In Section II-A, we have described the behaviour of the default kube-scheduler and at the end of Section II, we highlighted some of the limitations of this component. To address these limitations, we propose the integration of a robot installed on the master node, that can evaluate the hardware resources consumed by a running container and decide on what kind of nodes it can be deployed. So, to put our proposed solution into practice, developers will have to expose test endpoints for the most resource-intensive use cases in their applications before they are deployed on the cluster. Thus, before the deployment of containers on the cluster, the kube-scheduler performs the following operations:

- When the kube-scheduler receives a new container for deployment purposes on the cluster, it first deploys it in the sandbox environment (virtual machines running on the master node);
- Once the said application is deployed in the sandbox environment, the robot starts and simultaneously tests all the endpoints exposed by the container. The idea is to overload this application as much as possible and monitor the resources it requires when it is running at full speed;
- A score S is given to a container based on the resources it needs when running at maximum speed :
  - If $S \geq T_{appHGPU}$, then the kube-scheduler will consider only nodes having high GPU and high CPU capacity for the deployment of the new container;
  - If $T_{appLGPU} \leq S \leq T_{appHGPU}$, then the kube-scheduler will only consider nodes having low GPU and high CPU capacity for the new container deployment;
  - If $T_{appHCPU} \leq S \leq T_{appLGPU}$, then the kube-scheduler will only consider nodes having acceptable CPU and high GPU capacity for the new container deployment;
  - If $T_{appLCPU} \leq S \leq T_{appHCPU}$, then the kube-scheduler will only consider nodes having acceptable CPU and low GPU capacity for the new container deployment;
- After the score S is assigned to an application, its deployment is faster and easily done because only some nodes will be considered depending on that score;
- Otherwise, all the nodes likely to host a pod are overloaded, the new pod will be deployed on the worker in the higher level.

Since thousands of applications are running at the same time on edge nodes in a smart city, it is important not to over-use or under-use the nodes by choosing which application to run on what kind of node, so that this node will execute user requests with the highest possible speed compared to the other nodes on the network. The parameters that we used to evaluate the score of an application are RAM, CPU, GPU, network bandwidth between nodes's components and storage (HDD or SDD). Each parameter is assigned a weight W: then the weight for the amount of RAM used by an application is W_{RAM}, for the amount of CPU used is W_{CPU}, for storage memory is W_{Stor}, for the GPU is W_{GPU} and for bandwidth is W_{BW}. Also, we define N_i as the amount of hardware resource i consumed by a running application: so, let N_{RAM} be the amount of RAM used by a running application, N_{CPU} the amount of CPU used by the same running application, N_{Stor} the amount of data stored by that application during its execution, N_{GPU} the amount of GPU used and N_{BW} be the amount of bandwidth used by that application. Since the measurement units of each parameter differ, their values have to be normalized. Thus, the score S of a running application is given by:

$$S = \frac{W_{RAM} \times N_{RAM}}{max_{RAM}} + \frac{W_{CPU} \times N_{CPU}}{max_{CPU}} + \frac{W_{Stor} \times N_{Stor}}{max_{Stor}} + \frac{W_{GPU} \times N_{GPU}}{max_{GPU}} + \frac{W_{BW} \times N_{BW}}{max_{BW}} \quad (1)$$

where max_{RAM} is the highest value of RAM memory among all the nodes in the cluster, max_{CPU} is the highest value of CPU among all the nodes in the cluster, max_{Stor} is the highest

TABLE I. NOTATIONS

Notation	Explanation
CPU	Central Processing Unit
GPU	Graphical processing unit
RAM	Random Access Memory
N_i	Amount of hardware resource i consumed by a running application
$T_{appHGPU}$	Threshold value for an application to be launched on nodes having high GPU capacities and high CPU capacities
$T_{appLGPU}$	Threshold value for an application to be launched on nodes having low GPU capacities and high CPU capacities
$T_{appLCPU}$	Threshold value of an application to be launched on nodes having acceptable CPU capacities and low GPU capacities
$T_{appHCPU}$	Threshold value of an application to be launched on nodes having acceptable CPU capacities and high GPU capacities

value of storage memory among all the nodes in the cluster, max_{GPU} is the highest value of GPU among all the nodes in the cluster and max_{BW} is the highest value bandwidth among all the nodes in the cluster.

Algorithm III.1 shows the different stages executed by the Kube-Scheduler before new containers deployment on worker nodes.

Algorithm III.1: Kube-scheduler awareness

Input: *App*: a containerized application to deploy on the cluster

Begin
- Reception of the container (*App*) to deploy on the cluster ;
- Deploy container *App* in the sandbox environment on Master ;
- Evaluate the score S of container *App* using equation 1 ;
- **if** $S \geq T_{appHGPU}$ **then**
  - only consider nodes having high GPU and high CPU capacity for the deployment of *App*;
- **if** $T_{appLGPU} \leq S \leq T_{appHGPU}$ **then**
  - only consider nodes having low GPU and high CPU capacity for the deployment of *App* ;
- **if** $T_{appHCPU} \leq S \leq T_{appLGPU}$ **then**
  - only consider nodes having acceptable CPU and high GPU capacity for the deployment of *App* ;
- **if** $T_{appLCPU} \leq S \leq T_{appHCPU}$ **then**
  - only consider nodes having acceptable CPU and low GPU capacity for the deployment of *App* ;
- Deploy container *App* on the selected list of nodes ;

C. Load-balancing Awareness

In smart cities, end users send thousands of requests to the edge layer of an edge computing-based IoT network. Default load-balancing algorithms implemented on kube-proxy are userspace, iptable and ipvs, which are not suitable in the edge computing environment. In this section, we propose to use RAP protocol [14] where we integrate other metrics like storage and GPU in the nodes evaluation process considering that the number of nodes for the redirection of requests is reduced since custom applications are running on custom nodes. When a request arrives at a node, it performs the following operations:

- If the current node is running the target pod and has sufficient hardware resources, it executes locally the request and sends the response to the user;
- Otherwise, if it is not running the target pod, it sends the later to the node having the best score [14] and that is running the target pod. The redirection is done faster than in [14] because only a few nodes will be considered for the redirection.

So in this approach, user requests are performed mainly local on the node that receives the request, reducing the risk of latency during request redirection. In addition, in the case of redirection, the selection algorithm runs faster because the list of nodes to be considered is reduced since only some nodes run particular applications. In Fig. 2 the default kube-scheduler and HPA deploy applications on worker nodes without considering the minimum requirements of that application. For example, the best nodes that can run App1 are worker 3 and worker 4 but since they are already running an instance of App1, the next instance of App1 is deployed on worker 1 which will not process requests faster on App1 because of its hardware resources. Also, the default userspace algorithm implemented on kube-proxy component for load-balancing distributes requests evenly among the worker nodes. Thus, when worker 4 receives request number 6 to be executed by App1, it forwards it to worker 1 since App1 on worker 1 is only performing request number 3. So in addition to the delay between worker nodes 1 and 4 (7ms), you need to add the time required by worker 1 to perform requests on App1 which is not negligible.

In Fig. 3, pods are deployed on the best nodes that can perform user requests faster. For example, App1 is deployed on worker nodes 3 and 4. Also, user requests are executed locally most of the time, and are only redirected if the pod receiving the request is overloaded or if the node receiving the request is not running the target pod; for example, request number 4 on worker 4 to App2 (in orange) is redirected on worker 1. Therefore, as shown in Fig. 3, the HPA can delete the second instance of App2 on worker 1, and the two instances of App3 on worker 2 since they are not used; this will contribute to free resources on these nodes.

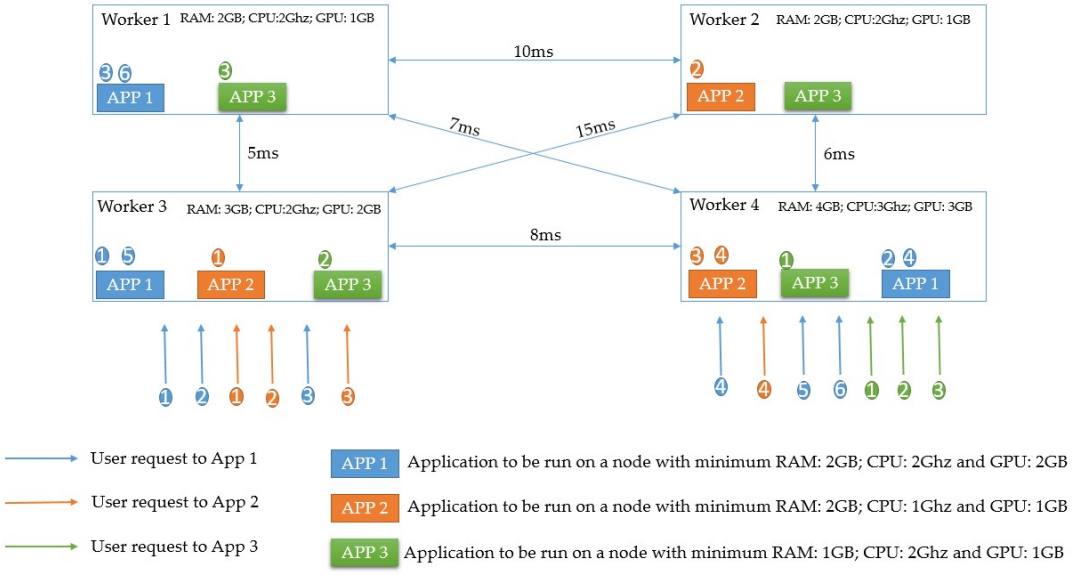


Fig. 2. Default kube-scheduler and userspace algorithm.

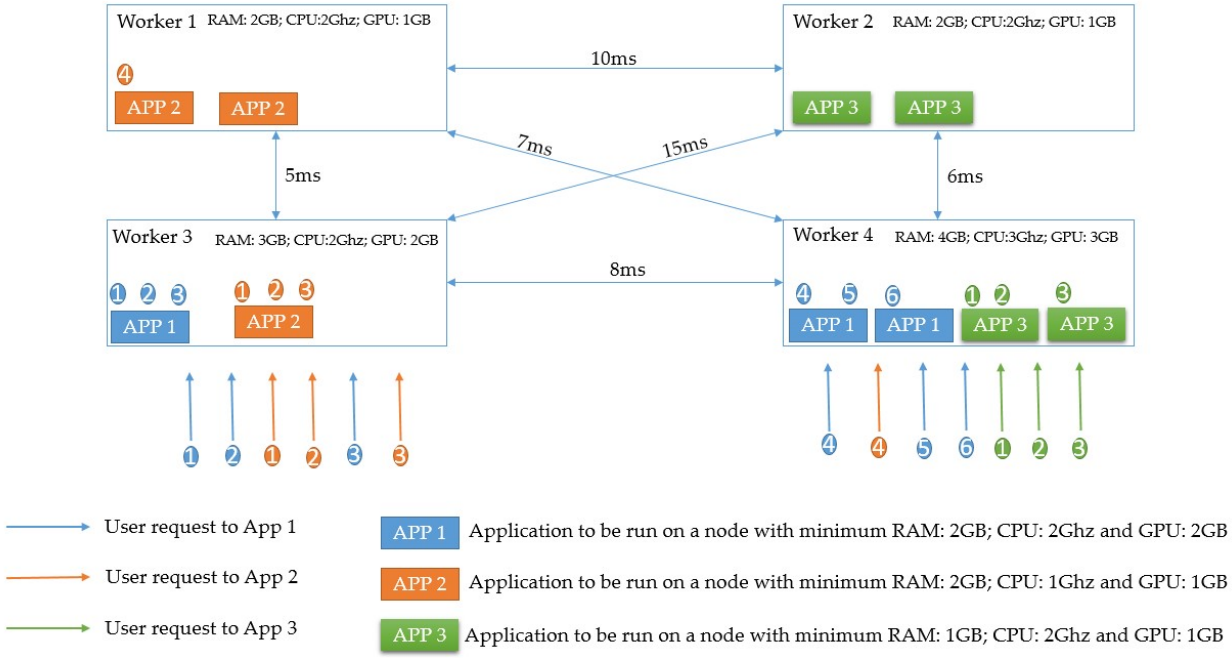


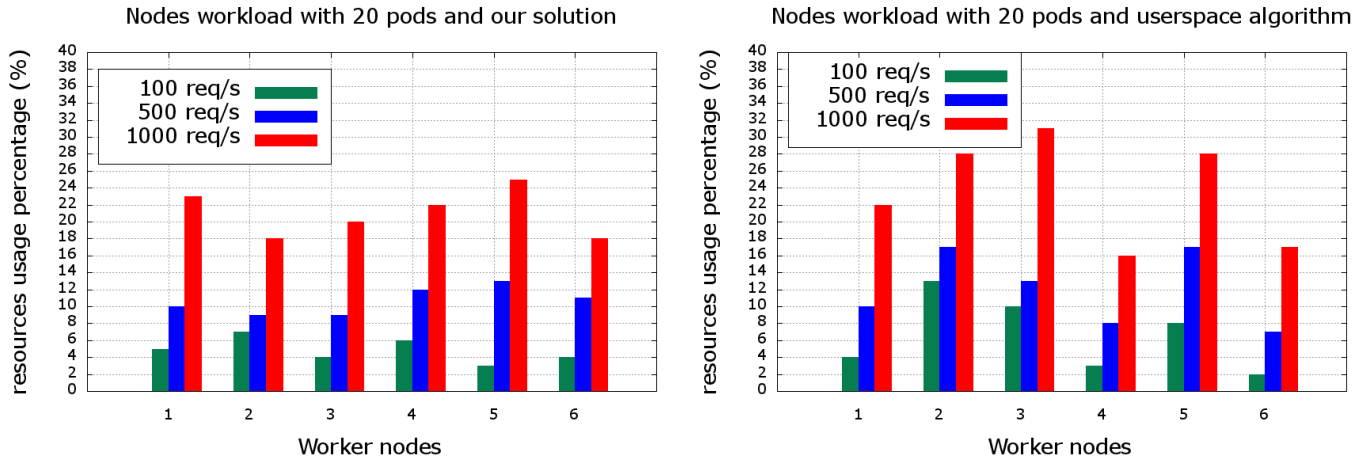
Fig. 3. Our extended kube-scheduler and load-balancing algorithm.

IV. Simulations and Results

In this section, we present the results of our simulations.We installed microk8s V1.26 which is a lightweight distribution of Kubernetes on a cluster with one master node and 6 worker nodes. The characteristics of each node are given in Table II. Also, requests to our Kubernetes cluster have been simulated with the Apache HTTP Server Benchmarking (AB) tool, which is a utility that tests the performance of a server. It has been designed to give an idea of the level of performance of an installation. In particular, it allows one to determine the number of requests an installation can process per second.

A. Workload on Worker Nodes

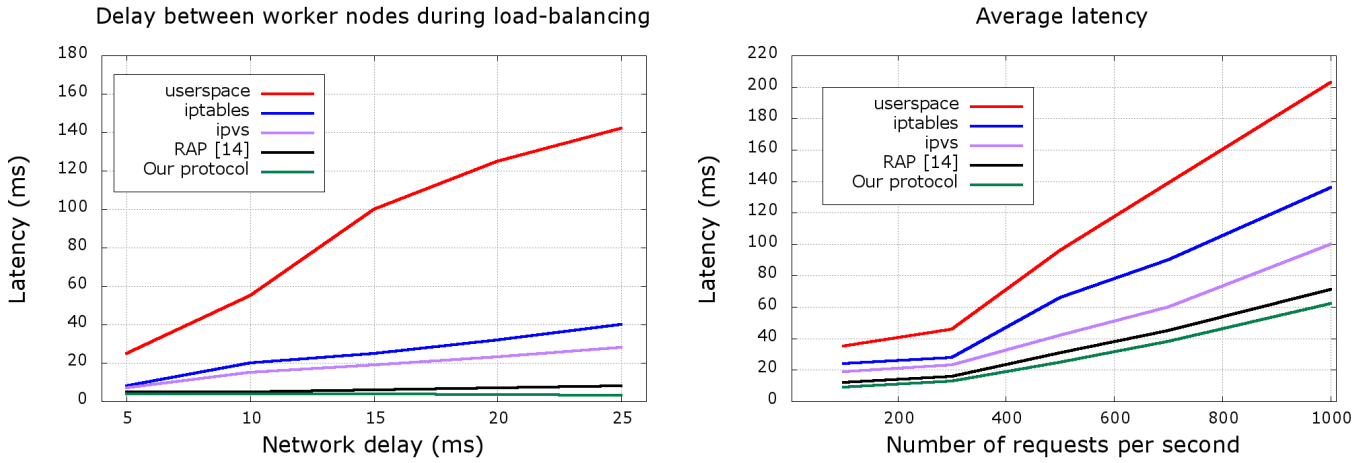
Fig. 4 shows the evaluation of the workload on each worker node when there are 100, 500 and 1000 requests per second (req/s) arriving at the cluster with 20 pods. Fig. 4a shows the results when nodes implement our load-balancing algorithm. All the pods are distributed among the worker nodes while considering their resources. The workload is well distributed among the nodes and there are no cases where some nodes are over-used while others are under-used. Fig 4a shows that when 500 req/s are arriving on the cluster, worker 1 uses 10% of its resources, worker 2 uses 9% of its resources, worker 3



(a) Workload on worker nodes with our approach

(b) Workload on worker nodes with userspace algorithm

Fig. 4. Evaluation of workload on worker nodes.



(a) Latency between worker nodes

(b) End users requests latency

Fig. 5. Latency evaluation.

TABLE II. NODES CHARACTERISTICS

Node	Hard drive type	Number of CPU	disc storage (GB)	RAM (GB)	CPU (Ghz)	GPU
Master	SSD	5	250	8	3.5	4
Worker 1	SSD	2	120	3	2.5	2
Worker 2	SSD	3	150	2	1.8	3
Worker 3	SSD	2	100	3	2.8	1
Worker 4	SSD	4	180	2	3.0	2
Worker 5	SSD	2	150	4	2.0	2
Worker 6	SSD	3	100	2	3.1	2

uses 9%, worker 4 uses 12%, worker 5 uses 13% and worker 6 uses 11% of its resources. However, Fig. 4b shows the results when the default userspace algorithm is implemented. It shows that, when 500 req/s are arriving on the cluster, worker 1 uses 10% of its resources, worker 2 uses 17% of its resources, worker 3 uses 13%, worker 4 uses 8%, worker 5 uses 17% and worker 6 uses 7%. Fig. 4b shows that, even if the requests are evenly distributed among the worker nodes, some pods are not deployed on the appropriate worker nodes; and therefore, these nodes are overused. These evaluations have been made while executing the most resource-intensive use cases of each pod. That is, with the use of our proposed solution, pods are deployed on the most suitable nodes in the cluster so that even if they are in demand, they can perform operations and deliver responses faster.

Results shown in Fig. 4 are supported by those presented in Table III, which shows the comparison between our protocol

TABLE III. EVALUATION OF STANDARD DEVIATION

Protocol	100 req/s	500 req/s	1000 req/s
Our protocol	1.37	1.49	2.59
userspace algorithm	3.97	4	5.73

and the default userspace algorithm in terms of the standard deviation of the workload on worker nodes. The standard deviation with our protocol is 1.37 when there are 100 req/s, while the one of the userspace algorithms is 3.97. When there are 500 req/s, the standard deviation with our protocol is 1.49 while the one of userspace is 4; and when there are 1000 req/s, the standard deviation with our protocol is 2.59 while the one of userspace is 5.73. These results show that the workload on worker nodes when using our protocol does not vary a lot compared to the cases with userspace.

B. Evaluation of Delay Between Worker Nodes During Load-balancing and Its Impact on Latency

Fig. 5a shows the evolution of the latency when the network delay between two worker nodes increases. Since in edge computing, worker nodes are geographically dispersed, the latency between worker nodes plays a key role in load-balancing management. Our protocol was compared to the RAP protocol [14] and the default routing decisions implemented in kube-proxy such as userspace, iptables and ipvs. Fig. 5a shows that userspace is the worst because of its round-robin algorithm and its performance varies from 25ms to 140ms; iptable which varies from 8ms to 40ms is better than userspace because while performing its random algorithm it often selects the closest node to transfer requests; ipvs which is tailored for large scale applications perform better than userspace and iptable. RAP protocol [14] is better than all the default routing decisions implemented in kube-proxy because it executes as many as possible requests locally on the node that received the request, and forwards them only if the local node has no capacity; moreover, the forwarding of a request takes into consideration the hardware resource on the other nodes and the delay between them. Our protocol is the best because our routing algorithm is not only based on the one proposed by [14], but also takes less time because the workload is reduced, thanks to the scheduler distributed pods while considering the scores and the nodes on which they can be deployed. Thus for load-balancing purposes, only a few worker nodes are considered in the routing algorithm.

We also evaluated the average latency of our solution when the cluster receives 100 to 1000 requests per second and compared the obtained results with userspace, iptables, ipvs and RAP [14] as presented in Fig. 5b. While latency with userspace varies from 35ms to 203ms, it varies from 24ms to 136ms in iptables, from 19ms to 100ms with ipvs, from 12ms to 71 ms in RAP, and finally from 9ms to 62ms in our solution. Our solution is the best because of two main reasons, the pods are distributed on suitable nodes which perform user requests faster and secondly, it is based on RAP protocol [14] for load-balancing algorithm. Fig. 4a shows that the workload is well distributed among the nodes even when several pods are running, and Fig. 5a shows that even if the delay between worker nodes increases, the latency for load-balancing management between those nodes is still weak. So, the fusion of these results helps to show that the network latency of our solution is the best.

V. CONCLUSION

This paper presents our approach to enhance latency in an edge computing-based IoT network with Kubernetes as the container orchestrator. This approach extends the default kube-scheduler component deploys pods on the best nodes in the cluster, and then distributes requests among the worker nodes to manage load-balancing. The results show that this approach is promising and better than other approaches in terms of latency. In future work, we plan to examine throughput and a politic for moving containers from one node to another when necessary to increase latency. We also plan to reduce latency by deploying pods according to the geographical regions in which they are used effectively, using machine learning techniques.

REFERENCES

[1] Hemant Kumar Apat, Rashmiranjan Nayak, and Bibhudatta Sahoo. A comprehensive review on internet of things application placement in fog computing environment. *Internet of Things*, 23:100866, 2023.

[2] The Kubernetes Authors. Kubernetes cluster architecture. https://kubernetes.io/docs/concepts/architecture/, 2024-04-22. 2023-10-23.

[3] Sebastian Böhm and Guido Wirtz. Profiling lightweight container platforms: Microk8s and k3s in comparison to kubernetes. In *ZEUS*, pages 65–73, 2021.

[4] Sebastian Böhm and Guido Wirtz. Cloud-edge orchestration for smart cities: A review of kubernetes-based orchestration architectures. *EAI Endorsed Transactions on Smart Cities*, 6(18), 2022.

[5] Ivan Cilic, Petar Krivic, Ivana Podnar Zarko, and Mario Kusek. Performance evaluation of container orchestration tools in edge computing environments. *Sensors*, 23(8), 2023.

[6] Ioannis Dimolitsas, Dimitrios Spatharakis, Dimitrios Dechouniotis, Anastasios Zafeiropoulos, and Symeon Papavassiliou. Multi-application hierarchical autoscaling for kubernetes edge clusters. In *2023 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 291–296, 2023.

[7] Raphael Eidenbenz, Yvonne-Anne Pignolet, and Alain Ryser. Latency-aware industrial fog application orchestration with kubernetes. In *2020 Fifth International Conference on Fog and Mobile Edge Computing (FMEC)*, pages 164–171, 2020.

[8] Tom Goethals, Filip De Turck, and Bruno Volckaert. Extending kubernetes clusters to low-resource edge devices using virtual kubelets. *IEEE Transactions on Cloud Computing*, 10(4):2623–2636, 2022.

[9] Tom Goethals, Bruno Volckaert, and Filip De Turck. Adaptive fog service placement for real-time topology changes in kubernetes clusters. In *CLOSER*, pages 161–170, 2020.

[10] Yiwen Han, Shihao Shen, Xiaofei Wang, Shiqiang Wang, and Victor C.M. Leung. Tailored learning-based scheduling for kubernetes-oriented edge-cloud system. In *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, pages 1–10, 2021.

[11] Rathinaraja Jeyaraj, Anandkumar Balasubramaniam, Ajay Kumara M.A., Nadra Guizani, and Anand Paul. Resource management in cloud and cloud-influenced technologies for internet of things applications. *ACM Comput. Surv.*, 55(12), mar 2023.

[12] Paridhika Kayal. Kubernetes in fog computing: Feasibility demonstration, limitations and improvement scope : Invited paper. In *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*, pages 1–6, 2020.

[13] Seong-Hyun Kim and Taehong Kim. Local scheduling in kubeedge-based edge computing environment. *Sensors*, 23(3), 2023.

[14] Quang-Minh Nguyen, Linh-An Phan, and Taehong Kim. Load-balancing of kubernetes-based edge computing infrastructure using resource adaptive proxy. *Sensors*, 22(8), 2022.

[15] Juan Marcelo Parra-Ullauri, Hari Madhukumar, Adrian-Cristian Nicolaescu, Xunzheng Zhang, Anderson Bravalheri, Rasheed Hussain, Xenofon Vasilakos, Reza Nejabati, and Dimitra Simeonidou. kubeflower: A privacy-preserving framework for kubernetes-based federated learning in cloud–edge environments. *Future Generation Computer Systems*, 157:558–572, 2024.

[16] Le Hoang Phuc, Linh-An Phan, and Taehong Kim. Traffic-aware horizontal pod autoscaler in kubernetes-based edge computing infrastructure. *IEEE Access*, 10:18966–18977, 2022.

[17] Yufei Qiao, Shihao Shen, Cheng Zhang, Wenyu Wang, Tie Qiu, and Xiaofei Wang. Edgeoptimizer: A programmable containerized scheduler of time-critical tasks in kubernetes-based edge-cloud clusters. *Future Generation Computer Systems*, 156:221–230, 2024.

[18] Samuel Rac and Mats Brorsson. Cost-effective scheduling for kubernetes in the edge-to-cloud continuum. In *2023 IEEE International Conference on Cloud Engineering (IC2E)*, pages 153–160, 2023.

[19] Rafael Vaño, Ignacio Lacalle, Piotr Sowiński, Raúl S-Julián, and Carlos E. Palau. Cloud-native workload orchestration at the edge: A deployment review and future directions. *Sensors*, 23(4), 2023.

[20] Cecil Wöbker, Andreas Seitz, Harald Mueller, and Bernd Bruegge. Fogernetes: Deployment and management of fog computing applications. In *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, pages 1–7, 2018.

Advanced Techniques for Optimizing Demand-Side Management in Microgrids Through Load-Based Strategies

Ramia Ouederni[1], Bechir Bouaziz[2], Faouzi Bacha[3]
Computer Laboratory for Electrical Systems, LR11ES26, INSAT, University of Carthage, Tunisia[1, 2, 3]
Electronics & Telecommunications Department, ISIMG, University of Gabes, Tunisia[2]
Electrical Department, ENSIT, University of Tunis, Tunisia[3]

***Abstract*—Microgrids are crucial for ensuring reliable electricity in remote areas, but integrating renewable sources like photovoltaic (PV) systems presents challenges due to supply intermittency and demand fluctuations. Demand-side management (DSM) addresses these issues by adjusting consumption patterns. This article explores a DSM strategy combining load shifting (shifting demand to periods of high PV generation), peak clipping (limiting maximum load), and valley filling (redistributing load during low-demand periods). Implemented in MATLAB and tested on a PV-battery microgrid, the strategy significantly reduces peak demand, improves the peak-to-average demand ratio (PAR), and enhances system stability and flexibility, particularly with the inclusion of deferrable loads.**

***Keywords*—Demand side management; microgrid; load shifting; peak clipping; valley fill**

I. Introduction

In recent decennials, an evolution in conventional power grid structure has taken place due to uncertain distributed renewable generation (URG) and advanced communication and control technologies [1], [2]. The increased utilization of deep penetration URGs and the improvement of demand-side resources (DSR) development have been seen as two important technological advances for supporting the sustainable expansion of today's energy infrastructure, enabled by the objective of total cost reduction [3]. A microgrid (MG) that can be incorporated with various URGs, advanced energy storage systems (ESS), and DSRs represents a sustainable way to deliver renewable, clean electricity as a vital part of the power transformation of conventional electricity production systems [4].

Microgrids (MGs) are stand-alone systems that support the integration of distributed power resources (DERs) and renewable energies for both reliability and profitability [4]. They are important components of intelligent energy transmission and distribution networks that regulate the electrical system's resilience under unfavorable conditions.

The various solutions for supplying rural areas with electricity are extensions of the national grid, autonomous systems, or microgrids [5]. Microgrids, or small electrical networks, have become a major solution because they can function independently. They are essential to developing an electrical infrastructure built around locally available renewable energy technologies at considerably reduced costs. In addition, they can accelerate access to electricity in areas where it is impossible to extend the grid in the short or long term.

To ensure that the electrical system becomes more reliable and robust, the peak demand is considered as opposed to the average demand. As a result, environmental resources are being wasted and production systems are being under-used. Rapid-response power generators (e.g. gas and oil-fired units), employed to satisfy the demand peak, not only cost a lot of money, they also generate high levels of carbon emissions. As a response, various packages were presented to help users shape their energy use profiles. These programs aim to make efficient use of the generation available so that new transmission and generation infrastructure is installed as little as possible. Such programs, referred to as demand-side management (DSM) programs, are designed to program consumption or reduce it [6].

A DSM program provides support for power grid functionalities in various areas, such as electricity market control, infrastructure maintenance, and management of decentralized energy resources [7]. In electricity markets, it informs the load controller about the latest load schedule and possible load reduction capabilities for each time step of the next day. Using this procedure, it schedules the load according to the objectives of interest associated with the power distribution systems [8], [9]. The load shapes indicate industrial or residential consumers' daily or seasonal electricity demands between peak hours (PHs) and off-peak hours (OPHs). These shapes can be modified by six techniques [9], [10]: peak clipping, valley filling, load shifting, strategic conservation, strategic load growth, and flexible load shape.

Both peak clipping and valley filling are techniques for controlling loads directly. While peak shaving is concerned with peak load reduction, valley filling takes into consideration load shaping on off-peak loads. The best and widely employed method of managing load in today's electricity networks is load shifting. It involves shifting the load from the PH to the OPH. In the case of conservation strategy [10], demand reduction methods are used at the customer's end to obtain optimal charging patterns. If the load requirement is greater, the daily response is optimized using distributed energy resource load augmentation techniques [11], [12].

The principal aims of the DSM are to remodel a load profile and regulate the highest and lowest day-to-day power requirements to make the most of power sources and prevent the need to construct new generation power plants. This may result in the consumption of electricity. Many DSM technologies are designed to maximize efficiency to avoid or reduce the need to build more power plants. Further reasons to implement DSM techniques include social relations and environmental considerations, by modifying customers' power consumption patterns at peak and off-peak times to conserve power [13].

In this paper, three DSM methods were applied for a microgrid of model photovoltaic solar panels with real measured demand as well as radiance data to show the benefits of energy efficiency which is currently lacking on the island of Djerba, Tunisia. A model system was simulated in the MATLAB software using the DSM load-carrying capability. The principal research contributions of this work are summarized in detail below.

- The DSM methods of load shifting, load capping, and valley filling were applied on a microgrid model of the island of Djerba, Tunisia.
- Substantial power consumption savings can be realized through corresponding generation and load demand requirements without deep-discharging of battery storage.
- The global load profile is enhanced by exploiting available renewable energy resources to a large extent, thus minimizing total reliance on the storage system.

The remaining sections in this paper are structured in the following way. In Section II, we describe the configuration of the microgrid, in Section III we present the DSM method and its features, and in Section IV we present the DSM problem. Section V presents the detailed results and discussion, while Section VI concludes the paper.

II. Configuration of a Microgrid System

A schematic representation of the proposed microgrid system is shown in Fig. 1. This system comprises three different groups of units. The first includes energy production units with photovoltaic panels. The second is comprised of energy consumption units, principally residential electrical loads. The third is dedicated to energy storage, as represented by a typical battery bank.

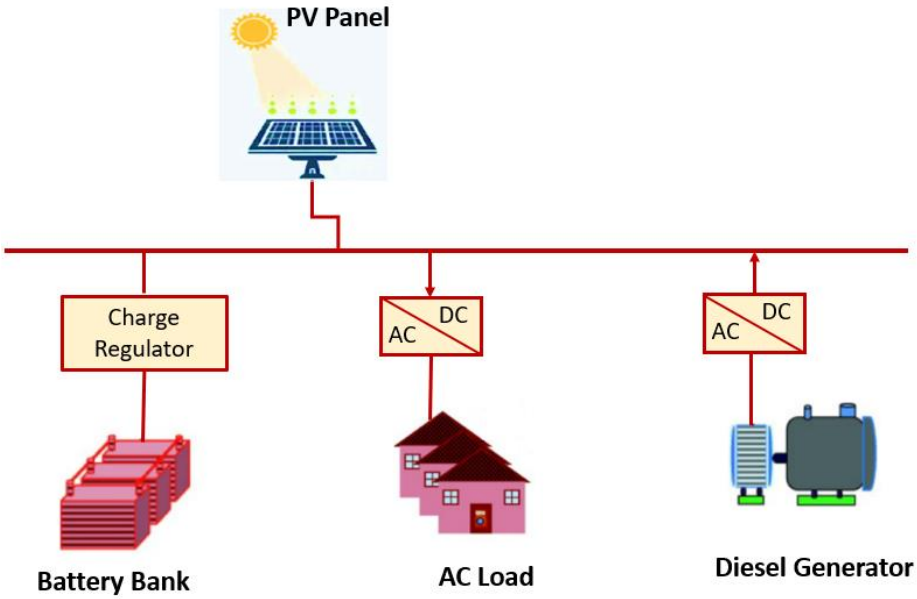


Fig. 1. Schematic representation of the proposed microgrid system.

III. Demand Side Management

Demand side management (DSM), also known as energy demand management, consists of implementing strategies and procedures to improve the efficiency of an energy system on the supply and demand side. It enables us to manage, control, influence, and generally minimize power consumption, giving us a better perspective on the potential of renewable energy systems, or making effective use of those systems that have been installed. Such strategies can provide substantial financial advantages to energy end-users and reduce carbon emissions. The measures aimed at reducing energy consumption do not necessarily have to be costly, and can often even be provided free of charge, making them easy to implement in development regions [14]. According to the [15] definition of demand-side management, DSM comprises three main concepts: demand response, energy efficiency, and energy conservation.

The demand response (DR) approach involves modifying customer electricity consumption, shifting load from periods of peak demand to off-peak times, and using active demand pricing strategies: by modifying electricity prices over time (when it's off-peak, it's less expensive) or through inducement payments, these offer customers an incentive to reduce their demand for energy at a particular period of the day. DR is usually linked to advanced energy metering technologies that provide information to end-users on their consumption of energy.

DSM can be implemented in several ways, and Fig. 2 shows several categorizations of DSM with associated load shapes. Peak clipping aims at reducing demand during peak hours. Utilities achieve this control by incentivizing customers not to consume power during peak hours, directly controlling loads, or setting up higher prices. The method is helpful in cases with no possibility of setting up or installing new power plants [16]. Valley filling focuses on raising usage during periods of very low electricity profile to keep demand and supply balanced, avoiding the startup and ramp-up costs of generators [17]. Load growth is more common when using electric vehicles, where customers are encouraged to increase usage up to a certain threshold for grid stability [18]. Load shifting gives consumers options to shift their usage pattern to off-peak hours based on cheap tariffs.

It is the combination of load clipping and valley filling [19], [20]. Flexible load shaping is when consumers are flexible enough to shift their loads to different low-usage slots. Usually, customers willing to participate in this are identified and incentivized for their participation [16]. Energy efficiency is when the overall load profile is lowered throughout the day by using more energy-efficient devices or through cyclic operation [21].

A demand curve is formed by various DSM technologies [21], as illustrated in Fig. 2, additionally, but without limitation:

- Load shifting: The main objective of load shifting is the transfer of load from peak to off-peak hours.
- Conservation: The best-known approach to reduce global consumption of electricity, and not only at peak periods.
- Peak clipping: The main aim of this strategy is to reduce demand at peak times (e.g. 7 p.m.). Energy consumption is reduced by the control of equipment that can be switched off, like air conditioners or heaters, by the customer or the electricity supplier.
- Valley fill: Under valley fill, the demand for load increases in off-peak hours and reduces during peak hours.
- Load growth: Unlike the conservation approach, with load expansion, the overall amount of electricity sold (off-peak and on-peak) increases.
- Load Shape Flexibility: With this technique, the utilities have the power to discontinue loading as required, and without notifying the customer. Changes in the quality or quantity of service are generally described as flexible load shapes.

Practically all the DSM programs are designed and implemented to optimize the use of existing power plants. Another crucial objective is to avoid, defer, or delay the construction of much-needed new power plants (both renewable and conventional).

A. Demand Response

These programs are recent developments in demand-side management, involving consumers' participation in improving their energy consumption patterns in response to instantaneous price variations [21]. They are seen as an effective solution to some of the problems of deregulated power systems.

Demand response enables industrial, commercial, and residential customers to optimize their electricity consumption to obtain reasonable prices and improve grid reliability. In other words, it can modify electrical energy consumption to reduce peak loads and shift consumption to off-peak hours [22]. Demand response methods generally fall into two categories: electricity price-based programs and incentive-based programs. Fig. 3 illustrates this classification.

In this article, among the various demand management programs, the load shifting program has been used to reduce peak load and increase load during periods of low demand (i.e. peak load reduction and load trough filling). There are also restrictions on time-of-use shifting for each of the loads.

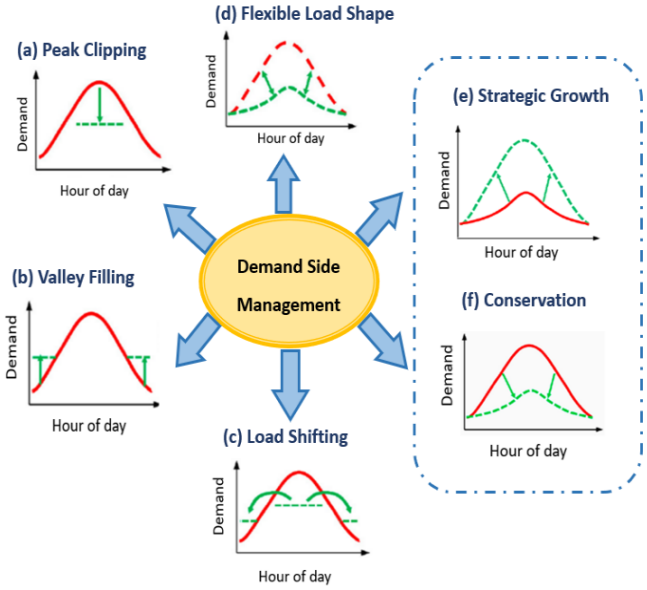


Fig. 2. Classification strategies for demand-side management.

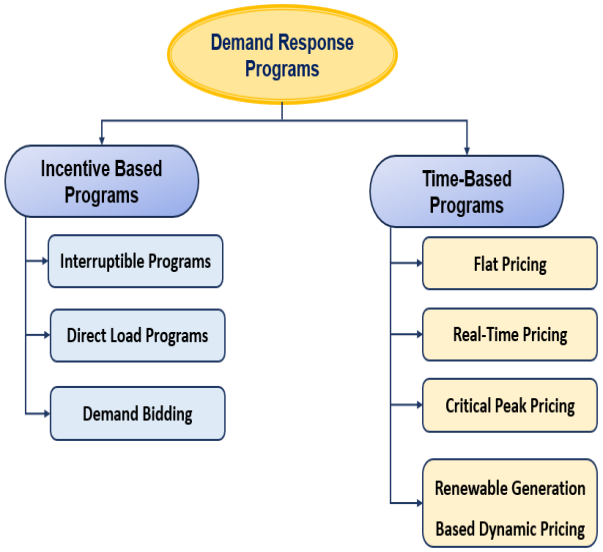


Fig. 3. Different types of demand response programs.

To illustrate the process of implementing the Demand Side Management (DSM) method, the following pseudo-algorithm describes the key steps in the approach used:

Algorithm 1: Pseudo-code of the DSM proposed algorithm

Step 1
Enter data, including renewable energy production (solar irradiance), load demands, system parameters (PV area, yield), and threshold values for demand-side management (DSM).
Step 2
Calculate hourly energy production using input data and system parameters.
Step 3
Initialize load curve for DSM operations (load shifting and peak shaving).
Step 4
Identify periods with significant energy production.
Determine the time of maximum energy production.
Step 5
For each period of the load curve:
If the load exceeds the DSM threshold:
Calculate excess load above a threshold.
Reduce the load to a threshold level.
Redistribute excess load to adjacent periods with sufficient energy production (Load Shifting).
Step 6
For each period of the modified load curve:
If the load still exceeds the threshold:
Limit load to threshold level using Peak Clipping.
Step 7
Apply the Valley Fill method:
Identify periods of low demand (valleys) in the load curve.
Redistribute non-critical load from high-demand periods (after Load Shifting and Peak Clipping) to these low-demand periods to smooth the load curve.
Step 8
Repeat steps 5, 6, and 7 until the load curve is optimized.
Step 9
Return the optimized load curve and analyze the results.
End

IV. Problematic Demand Management System

For a microgrid system that includes the use of photovoltaic (PV) generation, optimization of power consumption is a crucial factor in reducing costs and maximizing the utilization of renewable resources. Two strategies commonly used to meet this objective are demand-side management (DSM):

- Peak Clipping: The purpose of this strategy is to reduce the maximum power requirement by clipping and reducing loads above a specific level.
- Load Shifting: With this approach, loads are shifted during peak periods to periods when photovoltaic generation is most available, thus minimizing dependence on conventional sources.
- Valley Fill: In this method, the loads in periods when demand is low are increased to fill the valley of demand, thus leveling out demand and optimizing the use of available resources, thereby minimizing variations and dependence on traditional sources of energy.

A. System Modeling Equations

1) Power generated by the photovoltaic system: The output generated by the photovoltaic panels, P_{PV} (t), is given in the equation below:

$$P_{PV}(t) = \eta_{PV} \times A_{PV} \times I_{rr}(t) \quad (1)$$

Where:

η_{PV}: representing the efficiency of the photovoltaic panels.

A_{PV}: represents the overall area of the panels in m2.

$I_{rr}(t)$: represents the solar radiation in W/m2 at the time t.

2) Peak clipping strategy: The peak reduction is achieved by setting a maximal limit P_{seuil} above which the load output is restricted:

$$P_{load\_clipped}(t) = \min(P_{load}(t), P_{seuil}) \quad (2)$$

Where:

$P_{load}(t)$: is the initial load at the current time.

P_{seuil}: is the limit value selected.

3) Load shifting strategy: The load shift consists of shifting the surplus charge from the peak times to the times when the photovoltaic generation is highest. If at a given time t, the $P_{load}(t)$ load is above the limit, the $P_{exces}(t)$ surplus is determined as described below:

$$P_{exces}(t) = P_{load}(t) - P_{seuil} \quad (3)$$

This excess is allocated to the hours of most significant PV generation (defined as more than 50% of the maximum PV generation):

$$P_{load\_shifted}(h) = P_{load\_shifted}(h) + \frac{P_{exces}(t)}{N_{PV}} \quad (4)$$

Where:

h: denotes a time selected for redistribution.

N_{PV}: is the total number of hours selected for redistribution.

4) Valley-filling strategy: A valley fill approach involves the redistribution of under-utilized capacity at times of low demand (off-peak hours) to cover the valleys in the load curve. If the load $P_{load}(t)$ is less than a predetermined level at a given time t, the load deficit $P_{deficit}(t)$ is calculated in the following equation:

$$P_{deficit}(t) = P_{seuil} - P_{load}(t) \quad (5)$$

To smooth demand, this deficitis then assigned to the periods of lowest photovoltaic production (defined as less than 50% of peak photovoltaic output):

$$P_{load\_filled}(h) = P_{load\_filled}(h) + \frac{P_{deficit}(t)}{N_{low\_PV}} \quad (6)$$

Where:

h: represents a period selected for the redistribution.

N_{low_PV}: represents the number of total hours selected for the redistribution when the photovoltaic output is low.

5) Mathematical formulation of optimization: The general objective of the problem is to minimize the peak loads and maximize the utilization of renewable energy. The problem can be expressed in the following terms:

$$\min\left(\max P_{load\_final}(t)\right) \quad (7)$$

With the constraints below for peak clipping:

$$P_{load\_final}(t) \leq P_{seuil} \quad (8)$$

Excess load is redistributed in a balanced manner during load shifting, with consideration of PV generation. This approach ensures that load adjustments align with solar energy production, optimizing efficiency and reducing reliance on non-renewable sources

V. Results and Discussion

A. Basic Case

This study utilized a dataset from a microgrid located on the island of Djerba in Tunisia, with a maximum solar power capacity of 889.128 kW. This remote microgrid is designed to supply power to the island's communities, contributing to the region's energy independence and sustainability. The analysis was conducted over a week during which irradiance measurements were taken, to assess the load profile and irradiance patterns to evaluate the solar power generation of the system. Fig. 4 illustrates the power consumption over this period, Fig. 5 presents the irradiance data, and Fig. 6 represents the photovoltaic system's power output. Comparing these results with similar studies in microgrid systems reveals consistent trends in power production and consumption patterns, particularly in remote island contexts. Studies conducted in other island communities show similar variability in solar irradiance and load profiles. However, our findings suggest that the Djerba microgrid demonstrates higher efficiency in converting solar energy, likely due to optimized PV system design and favorable climatic conditions. This has important consequences for isolated microgrid design and operations, underlining the necessity for an enhanced energy control strategy. In comparison to earlier research, this study underscores the potential for enhancing load-shifting techniques and optimizing storage systems, which could further reduce reliance on backup generators and improve overall energy efficiency in isolated microgrids.

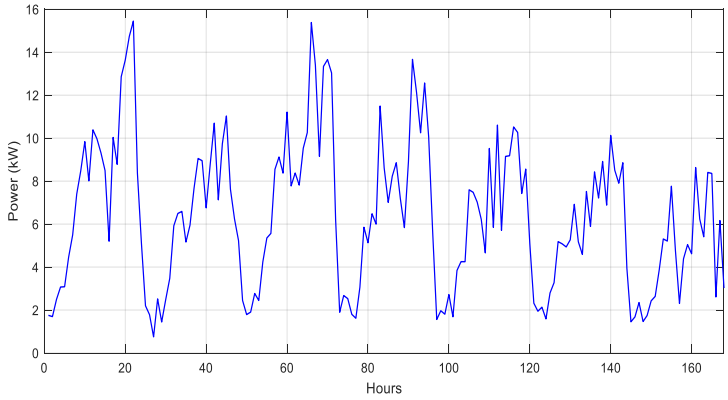


Fig. 4. Weekly load power analysis.

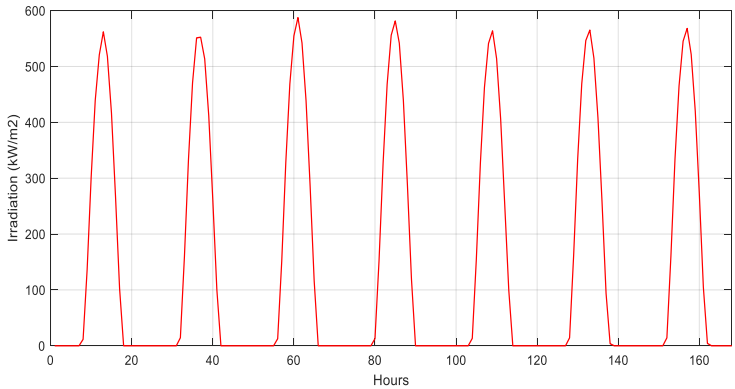


Fig. 5. Weekly analysis of solar irradiation levels.

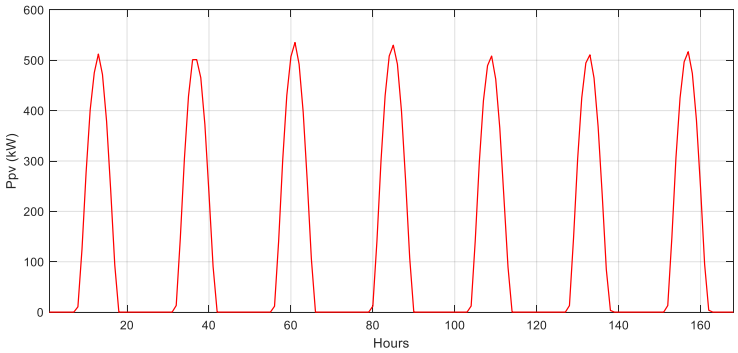


Fig. 6. Weekly performance analysis of photovoltaic power system.

B. *Effect of Peak Clipping*

A controller or switching equipment carries out the peak clipping process. The consumption is continuously supervised and when it rises to levels that give reason for concern, the controller can switch off specific loads to minimize the demand. The loads that can be disconnected are ranked in order of priority, and classified as transferable and non-transferable.

Fig. 7 illustrates the impact that peak clipping has on the hourly load. The blue curve shows the initial load, with significant daily fluctuations. The second curve represents the load after peak shaving, with load values restricted to a 10-kW threshold. The peak clipping process limits the peak consumption to a specified threshold, thereby preventing system surges and stabilizing the energy demand. It effectively holds the load below a monitored level, thus minimizing the potential for power system instability and overloading.

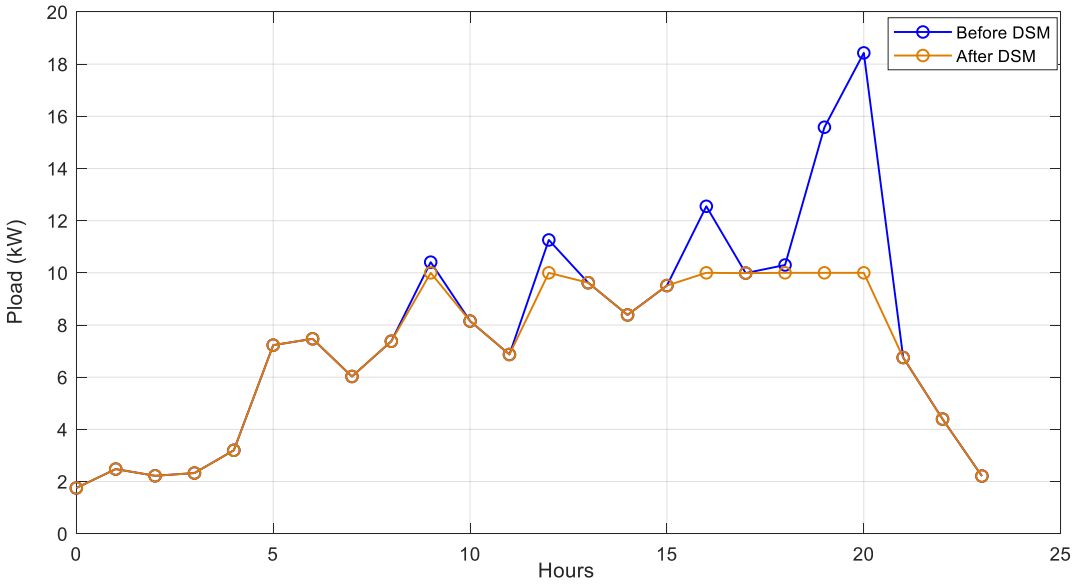


Fig. 7. Impact of peak clipping DSM on load profile.

C. *Effect of Load Shifting*

Fig. 8 presents an example of demand-side management (DSM) implementation through load shifting. An analysis of the results indicates that load shifting has been strategically applied over 24 hours, with most operations scheduled during peak solar irradiation to minimize the need for energy storage. By aligning load utilization with times of highest production, the system maintains a sufficient margin to supply critical loads during peak demand periods, thereby enhancing system stability and preventing overloads. The impact of load shifting on hourly load is illustrated in Fig. 8. The initial load, shown in blue, reveals significant peaks during high-demand periods, while the second curve represents the load after implementing load shifting, where consumption peaks have been effectively smoothed out. This method redistributes surplus load to hours of higher photovoltaic production, resulting in a more uniform demand distribution. Comparative analysis with similar studies shows that load shifting is a widely recognized strategy for optimizing energy consumption in microgrids, demonstrating its effectiveness in reducing peak loads and alleviating pressure on the power supply system. Our findings corroborate previous research, highlighting that load shifting not only enhances system reliability but also supports sustainability efforts by maximizing the utilization of renewable energy resources.

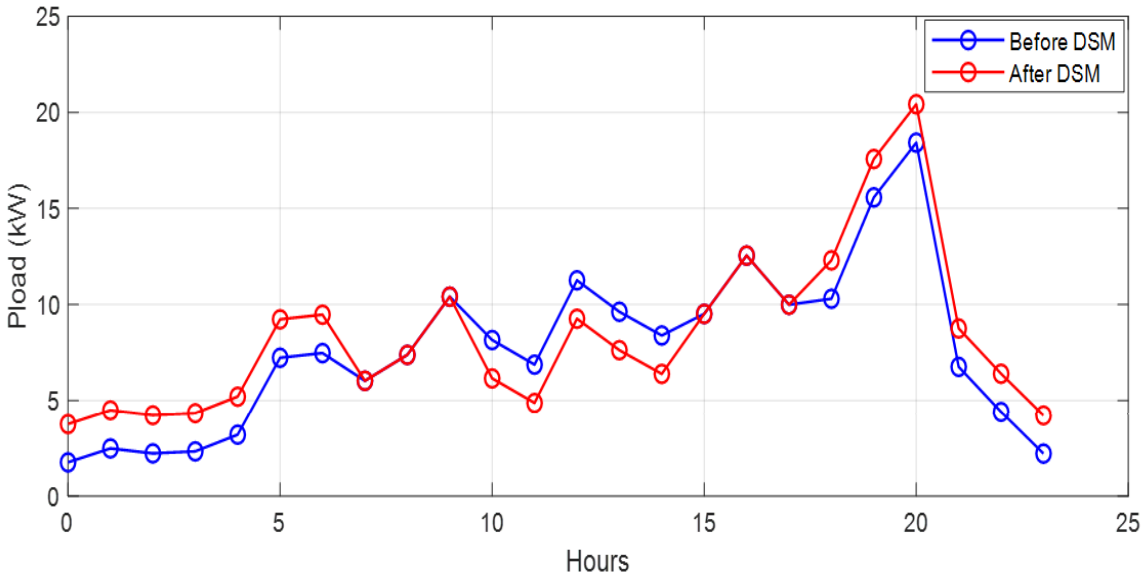


Fig. 8. Impact of Load Shifting DSM on 24-Hour Load Profile.

D. *Effect of Valley Fill*

The Valley Fill approach is implemented in the form of an intentional increase in load during times of reduced demand, usually monitored by an appropriate controller or energy management system. As consumption falls to a critical level, a controller may enable supplementary loads, prioritized based on their ability to transfer loads without affecting the global system. The loads that can be shifted are increased to compensate for any dips in the power curve.

Fig. 9 shows the effect of the Valley Fill method on the load per hour. The initial blue curve represents the load, characterized by steep peaks and troughs during certain periods of the day. The other curve shows the load after Valley Fill has been applied, with the troughs filled to smooth demand. This stabilizes power demand by preventing excess fluctuations, thus reducing the risk of under-utilizing the available energy resources and improving global system efficiency.

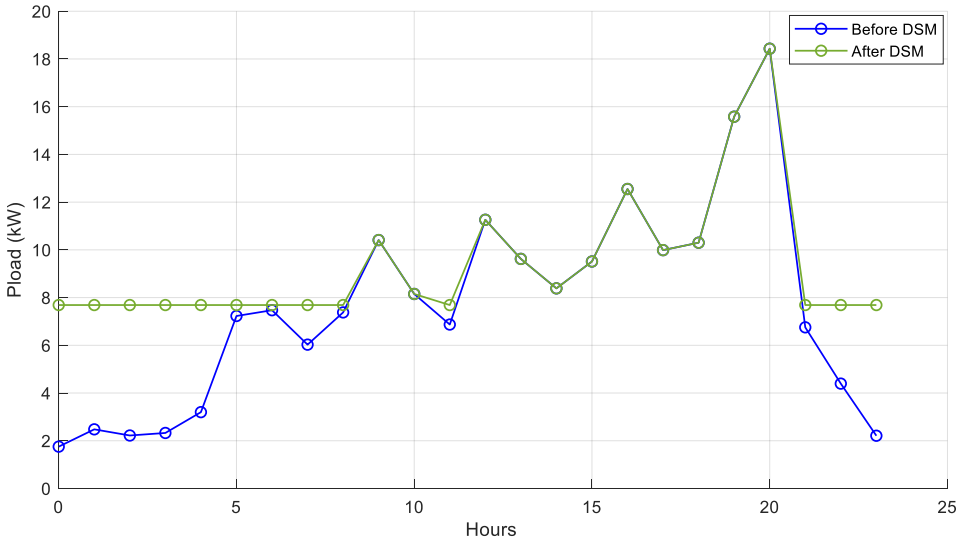


Fig. 9. Load Profile Analysis Before and After Valley Fill DSM.

Shifting reduces power demand and inconvenience in peak hours by switching surplus demand away from the system. Load shifting enables off-peak hours to be rescheduled with a smaller energy consumption reduction, as illustrated in Fig. 10.

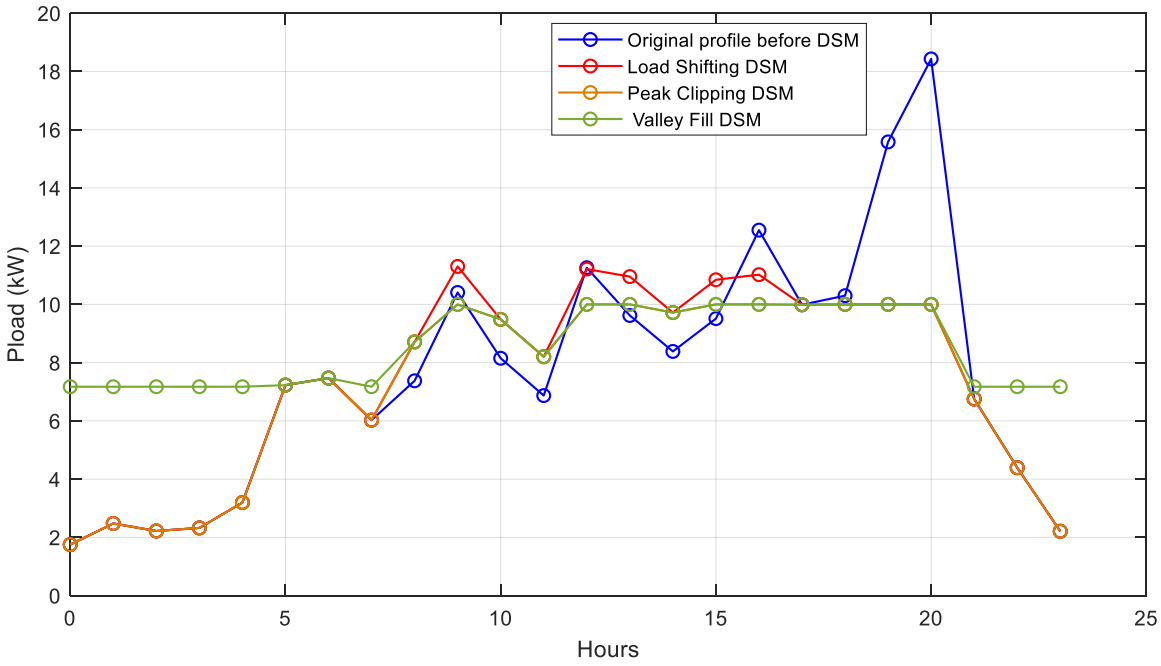


Fig. 10. The impact of DSM strategies on load profile: load shifting, peak clipping, and valley fill.

VI. Conclusion

This study examines energy management in a solar-powered microgrid using the Demand Side Management (DSM) method. The DSM approach was applied to a 24-hour solar microgrid on the island of Djerba, Tunisia, utilizing data from NASA. The results show that after implementing DSM strategies, the load profile aligned more closely with solar generation, as much of the energy demand was shifted to periods of higher solar output. This approach provides a framework for achieving more resilient isolated microgrids by optimizing the use of renewable energy. Additionally, reducing energy consumption in the microgrid enhances system stability and offers cost savings. By improving the balance between load and generation, the method also reduces the need for extensive storage capacity, especially during periods of low solar power availability.

References

[1] S.Onika, S. Peyghami, A. Sangwongwanich, F. Blaabjerg, "Reliability aspects in microgrid design and planning: Status and power electronics-induced challenges, Renewable and Sustainable Energy Reviews," vol.159, pp.1364–0321,2022.

[2] Bouaziz, B., Bacha, F. Direct power decoupling control using optimal switching table combined with virtual flux estimator for PWM rectifier. Electr Eng 106, 1605–1617 (2024). https://doi.org/10.1007/s00202-023-02179-6

[3] F.Yang, and X. Xia,"Techno-economic and environmental optimization of a household photovoltaic-battery hybrid power system within demand side management," Renew. Energy 108, 132–143,2017.

[4] A.Maleki, Z. Eskandar Filabi, M.ANazari, "Techno-Economic Analysis and Optimization of an Off-Grid Hybrid Photovoltaic–Diesel–Battery System: Effect of Solar Tracker," Sustainability, vol.14, 7296,2022.

[5] M.Usman, M.Khan, A.Rana, "Techno-economic analysis of hybrid solar-diesel-grid connected power generation system, " Journal of Electrical Systems and Information Technology,2017.

[6] S. Nanibabu, S. Baskaran, P. Marimuthu, "Realization of Demand Side Management in Smart Grid," Optimization Techniques for Hybrid Power Systems, 378–406, 2024.

[7] J. Yang, W. Xu, K. Ma, J. Chen, W. Guo, "Integrated demand-side management for multi-energy system based on non-cooperative game and multi-energy pricing," Sustainable Energy, Grids and Networks, vol.34, 2352–4677,2023.

[8] R. Ouederni, B. Bouaziz and F. Bacha, "Design and evaluation of an island's hybrid renewable energy system in Tunisia," 2022 5th International Conference on Advanced Systems and Emergent Technologies (IC_ASET), Hammamet, Tunisia, 2022, pp. 418-423,2022.

[9] P.D.Necoechea-Porras, A.López, J.C. Salazar-Elena, "Deregulation in the Energy Sector and Its Economic Effects on the Power Sector: A Literature Review," Sustainability, vol. 13, 3429,2021.

[10] A.Shewale, A.Mokhade, N.Funde, N.D.Bokde, "A Survey of Efficient Demand-Side Management Techniques for the Residential Appliance Scheduling Problem in Smart Homes," Energies, vol.15, 2863,2022.

[11] B. Williams, D. Bishop, P.Gallardo, J.G. Chase, "Demand Side Management in Industrial, Commercial, and Residential Sectors: A Review of Constraints and Considerations," Energies, vol.16, 5155,2023.

[12] R.Ouderni, B.Bouaziz, and F. Bacha, "A Case Study of Hybrid Renewable Energy System Optimization for an Island Community based on Particle Swarm Optimization," Engineering, Technology & Applied Science Research. vol.14, 3, 14367–14373,2024.

[13] S. N. Bragagnolo, J. C. Vaschetti, and F. Magnago, "A technical and economic criteria comparison on demand side management with multi-level optimization model," IEEE Latin America Transactions 19(9): 1494–1501,2021.

[14] R. Ouederni, B. Bouaziz, F. Bacha, " Modeling and cost optimization of an islanded virtual power plant: A case study of Tunisia," Turk J ElectrPower Energy Syst, 2(2), 168–179, 2022.

[15] MA.Bhaagat, NH.Mirjat, SA.Khatri, A.Mahar, MA.Raza, Demand Side Management for the Energy-Efficient Future of Pakistan Using a Low Emission Analysis Platform, Engineering Proceedings, vol. 46(1):28,2023.

[16] S. Panda, S. Mohanty, P.K.Rout, B.K.Sahu, M.Bajaj, H.M.Zawbaa, S.Kamel, "Residential Demand Side Management Model, Optimization and Future Perspective: A Review," Energy, vol.8, 3727–3766,2022.

[17] S.Mujeeb, TA.Alghamdi, S.Ullah, A.Fatima, N.Javaid, T.Saba, "Exploiting Deep Learning for Wind Power Forecasting Based on Big Data Analytics," Applied Sciences, vol. 9(20):4417,2019.

[18] S. Maharjan, Y. Zhang, S. Gjessing, D.H.K. Tsang, "User-Centric Demand Response Management in the Smart Grid with Multiple Providers," IEEE Trans. Emerg. Top. Comput, vol.5, 494–505,2017.

[19] G.H.Philipo, Y.A.Chande Jande, T.Kivevele, "Demand-Side Management of Solar Microgrid Operation: Effect of Time-of-Use Pricing and Incentives," J. Renew. Energy 2020, 6956214,2020.

[20] R. Ouederni, B. Bouaziz, F. Bacha, "Optimization of a Hybrid Renewable Energy System Based on Meta-Heuristic Optimization Algorithms," International Journal of Advanced Computer Science and Applications, vol.15(7), 2024.

[21] P.Palensky, D. Dietrich, "Demand Side Management: Demand Response, Intelligent Energy Systems, and Smart Loads," IEEE Trans. Ind. Inform, 7, 381–388,2011.

[22] S.Feuerriegel, N.Dirk, "Integration scenarios of Demand Response into electricity markets: Load shifting, financial savings and policy implications," Energy Policy vol. 96, 231–240,2016.

Ensemble of Weighted Code Mixed Feature Engineering and Machine Learning-Based Multiclass Classification for Enhanced Opinion Mining on Unstructured Data

Ruchi Sharma[1], Dr. Pravin Shrinath[2]
Department of Information Technology,
SVKM's NMIMS, Mukesh Patel School of Technology Management & Engineering, Mumbai, India[1]
Associate Professor, Computer Engineering Department, SVKM's NMIMS,
Mukesh Patel School of Technology Management & Engineering, Mumbai, India[2]

***Abstract*—There is an exponential growth of opinions on online platforms, and the rapid rise in communication technologies generates a significant need to analyze opinions in online social networks (OSN). However, these opinions are unstructured, rendering knowledge extraction from opinions complex and challenging to implement. Although existing opinions mining systems are applied in several applications, limited research is available to handle code-mixed opinions of a non-structured nature where there is a switching of lexicons in languages within a single opinion structure. The challenge lies in interpreting complex opinions in multimedia networks owing to their unstructured nature, volume, and lexical structure. This paper presents a novel ensemble approach using machine learning and natural language processing to interpret code mixed opinions efficiently. Firstly, the opinions are extracted from the input corpus and preprocessed using proposed Extended Feature Vectors (EFV). Subsequently, the opinion mining system is implemented using a novel approach using weighted code mixed opinion mining framework (WCM-OMF) for multiclass classification. The proposed WCM-OMF model achieves an accuracy of 79.11% and 72% for the benchmark datasets, which is a significant improvement over existing Hierarchical LSTM, Random Forest, and SVM models and state-of-the-art-methods. The proposed solution can be implemented in opinion detection of other business sectors beneficial in obtaining actionable insights for efficient decision-making in enterprises and Business Intelligence (BI).**

Keywords*—Opinion mining; Machine learning; weighted ensemble; code mixed; Natural Language Processing; Business Intelligence; Online Social Networks

I. Introduction

With the advent of information and communication technologies, there is an exponential volume of user-generated content [1] available online. However, this data is primarily in the unstructured format in the form of opinions [2-4], suggestions, and forum posts. A further significant chunk of opinions generated are code-mixed. Code mixing is the alternating between two or more languages in a single opinion text source. This renders opinion mining complex to implement and interpret.

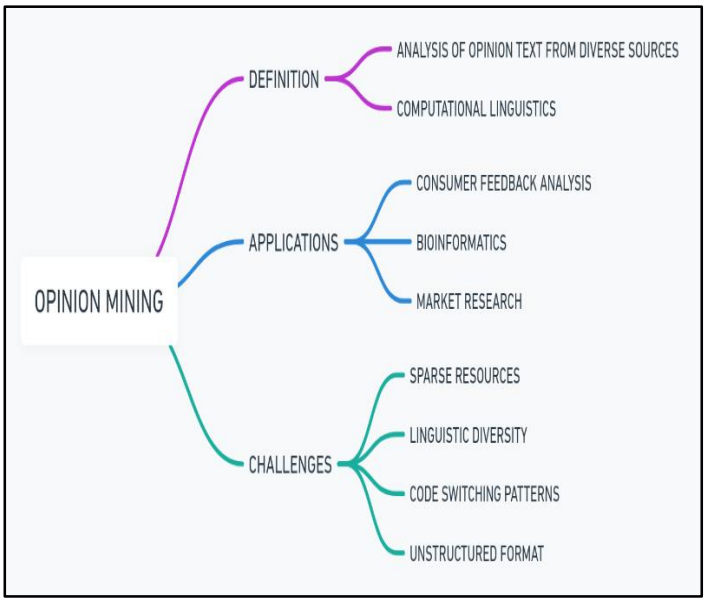


Fig. 1. Opinion mining trends and challenges.

Fig. 1 illustrates the opinion mining trends and challenges in various industrial sectors. The research specifically benefits multilingual cultures [5, 6] wherein multiple users convey the same expression differently despite belonging to the same regional area. Although there are studies related to opinion mining and opinion extraction, there is limited research on handling code-switching issues in opinion corpus. Also, existing studies focus more on mono-lingual input and overlook the lexical nuances of code mixed opinions. Given the aforementioned issue, we have developed a novel weighted code mixed opinion mining framework (WCM-OMF) for handling the code switching challenge. The proposed solution is the design and development of an ensemble method combining natural language processing and machine learning techniques. Compared to the existing systems, the proposed method achieves better results, improving model accuracy and F1 score metrics.

A. Applications of Opinion Mining

Opinion Mining is essential for achieving Business Intelligence in enterprises of several sectors as described in this subsection:

1) Consumer feedback analysis is one of the primary applications of opinion mining [7]. The extraction of consumer opinions and feedback is significant for enhancing the quality of products and services.

2) Manufacturing: Identifying opinions towards a product or service during a specific timeframe is highly valuable for firms, demonstrating the practical benefits of opinion mining in manufacturing.

3) Political analysis: Opinion mining of users about a political candidate [9, 10] or government policy can aid in understanding the public stance on a topic.

4) Risk analysis: Effective opinion mining can help prevent or mitigate risks and significant disasters in public sector companies.

B. Challenges of Opinion Detection

Existing Opinion Detection systems primarily focus on monolingual input. However, there is a vast amount of data in form of code-mixed lexicons. Detecting these opinions poses several challenges, as listed below:

1) Code-switch: The process of alternating between different lexicons in the intrasentential text [5] or opinion renders opinion detection and analysis complex to implement.

2) Lack of standardized datasets: The lack of standardized datasets [6] for machine learning used for interpreting code-mixed data can lower the performance of machine learning models.

3) Unstructured format: Online users utilize non-standardized formats while conveying information [2, 7] that can include abbreviations, special characters, commas, slang, and lexical resources not included in standard machine learning libraries required for interpreting these opinions. This non-colloquial opinion usage makes opinion interpretation complex to implement.

4) Low resource lexicons: Certain opinion corpus categories have low resource availability [9], especially for the lexicon database of languages that are scarcely available.

5) Healthcare domain: Applications in Bioinformatics [8] include techniques for analyzing unstructured clinical data and extracting patient feedback. It can lead to significant insights, especially those that can be overlooked by clinical practitioners.

6) Dialects: Opinions cannot be distinguished from one dialect [10] to another within the same language.

Table I illustrates sample instances of code mixed opinions, translated opinions using API services, and actual interpretation of the Opinion. As observed, there is a variation in actual meaning through accurate interpretation compared to direct translation owing to lexical nuances of code mixed opinions. In order to overcome this challenge, we propose the design and implementation of a Code mixed opinion mining framework using an Adaptive Lemma-based weighted approach and machine learning for multi-class classification.

TABLE I. SAMPLE CODE MIXED OPINION CORPUS

Sr No	Code Mixed Opinion	Translated Opinion	Actual Opinion Interpretation
O1	Online parksha toh ache hain but yaar cheating ko rokna thoda sa mushkil ho jaata hai	Online parking are very good, but it is a bit difficult to stop cheating	Online exams are good friend, but stopping cheating becomes a bit difficult
O2	Ningal ivide thanne irikku	You stay here	Please remain here
O3	Woh kal se gym ja raha hai gr8	He is going to the gym from tomorrow gr8	He's starting gym tomorrow, great.
O4	Enikku oru sambhavam undu parayan.	I have an event to tell.	I have a story to tell you

C. Research Contributions

The major contributions of this research are highlighted below:

1) Design and implementation of novel weighted code mixed algorithm for feature engineering in opinion mining.

2) Efficient handling of code switching problems in code mixed data input.

3) Implementation of multiclass classification using an ensemble of machine learning and natural language processing techniques.

4) We conduct a rigorous performance analysis with state-of-the-art techniques on benchmark datasets to compare the proposed model efficacy for code mixed opinions.

The rest of this research is structured as follows: In Section II, we discuss the background of the literature. Section III describes the proposed solution architecture. The results of the performance analysis and the comparative analysis of related studies are compared in Section IV. Discussion and conclusion are given in Section V and Section VI respectively.

II. LITERATURE REVIEW

In this section, we review the relevant literature and existing opinion mining systems related to our study and research objective. The Intelligent Multi Lexicon Opinion Mining (IMLOM) implementation taxonomy is summarized below and categorized as machine learning, deep learning, and hybrid techniques.

A. Supervised IMLOM Learning

For supervised IMLOM, researchers have adopted conventional machine learning techniques, namely Support Vector Machine (SVM), Naïve Bayes (NB), and Logistic Regression [11], for the Algerian social platform. Other studies also use machine learning approaches. Multimedia network

studies use ML [12] for various domains, like the work done on the climate change crisis [13-14], which highlights the priority levels of different region users and their regional dialects.

B. Unsupervised IMLOM Learning

For unsupervised IMLOM studies, it is observed that authors use deep learning in various application domains. For instance, medical domain analysis for gaining perspective regarding vaccinations [15] being mandatory post-COVID through a survey questionnaire on a sample population selected.

In study [16], deep clustering with embedding was used for document vector recommendations. Deep learning techniques have been applied in further studies with promising results [17-19].

Other studies used Graph Neural Network (GNN) [20], Long Short Term Memory (LSTM) [21], and stacked layering of LSTM. LSTM and stacking mechanism are utilized for the deep learning phase, combining machine and manual translation. Researchers have also employed Convolution Neural Network (CNN) [22] with efficient model development and lingual error analysis.

C. Ensemble IMLOM Learning

Various approaches are adopted for hybridization with variations in preprocessing and classification combinations. In [23], authors use sentence-level embedding combined with the Transformer approach. In study [24], authors hybridize bio-inspired optimization algorithms with recurrent neural networks (RNN) for political corpus. Other popular combinations include CNN, LSTM, Bidirectional LSTM (BiLSTM) feature extraction, and transformer approach [25-29]. Both quantitative and qualitative analyses are conducted to synthesize data effectively. The authors in [30] use the augmentation technique to perform further the LDA-based topic modeling approach and Multilingual bidirectional and Auto-Regressive transformers (mBART), achieving a 71% accuracy level and highlighting the challenge of handling informal text.

Researchers in study [31] achieved significant results by employing the traditional J48 machine learning approach on self-developed data from the extracted contents, applying several linguistic features to analyze the model. Study in [32] implements deep learning methods, and dual datasets are used for cross-pair mapping. However, testing a model trained on one system in another domain yields lower accuracy using source-to-target language mapping. In study [33], the authors highlight the need for a training corpus for Dravidian language content classification into multiple classes. The authors in [34] demonstrated target entity extraction successfully through decisions based on neighboring information. However, the performance was limited by the capabilities of CNN. Alternatively, the shallow neural network [35] is for Spanish input and achieves an F1 score of 0.48. In study [36], the authors preprocess the data corpus and then generate the vectorized using the pre-trained model for emotion detection in online text. Efficiency was lower for heterogeneous content than for the standard English language. Studies in [32, 37 38, 41] utilize LSTM and combinations of other deep learning techniques, including CNN and BiLSTM. Total of five languages, with approximately. 42K social media opinions [39]. The various contexts of words are tackled using a stacked version of the word embedding implementation [40]. Also, a discussion of several research groups currently working on this topic of offensive text in multilingual settings is highlighted [41]. Researchers in [42] proposed a novel method using a combination of deep learning techniques in a stacked format and compared it to standardized machine learning methodologies, including SVM, KNN, and logistic regression. The authors in [43] use Artificial Intelligence for heterogeneous datasets, implementing machine translation using the BART method on Hungarian European content. The study in [44] classifies opinions, employing deep learning techniques to explore opinions towards access to education. We investigate the research papers, and after rigorous analysis of the paper title, abstract, methodology, and conclusion, the initial list is filtered to the most relevant studies.

It is imperative to implement efficient opinion mining frameworks for handling linguistic variability, code-switching, and unstructured format to leverage this vast digital information available, code mixed in the unstructured format, enabling accurate decision making.

III. Proposed Methodology

This section presents the benchmark datasets extracted, data statistics, and the proposed workflow.

A. Dataset Statistics

This section describes the datasets utilized for the experimental evaluation of the proposed work, as explained in Table II. Here, we employ two Benchmark datasets, Bohra et al. [45] and Joshi et al. [46], specifically curated for the problem statement identified as having Filtered data points 4575 and 3879, respectively. These benchmark datasets from the ACL data repository have been further validated through quantitative validation using the Cohens Kappa ITA score. The Bohra dataset has a Kappa score of 0.92, and the Joshi dataset score is 0.64, making it suitable for the proposed framework statistical and qualitative performance validation.

TABLE II. Dataset Statistics for Experimentation

Statistic	Bohra Dataset	Joshi Dataset
Total Data Points Collected	112718	4981
Filtered Data Points	4575	3879
Positive Class Data Points	1661 (Offensive Speech Opinions)	1358
Neutral Class Data Points	Not Available	1939
Negative Class Data Points	2914 (Normal Speech Opinions)	582
Inter Annotator Agreement (Cohen's Kappa)	0.982	0.64

B. Proposed Opinion Mining Framework

In this section, the design and implementation of the proposed technique WCM-OM are explained in detail. Through an extensive literature analysis, we present the comprehensive Intelligent Code Mixed Opinion Mining framework. The majority of the existing frameworks rely solely on a single language or medium, resulting in a significant loss of relevant information. The proposed framework implements a fusion ensemble of statistical features and linguistic features of code

mixed opinions from the input corpus database. Also, we implement data preprocessing for code-switched data to effectively handle the issue of frequent code-switching in an intra-sentential opinion. Fig. 2 illustrates the conceptual framework of the proposed framework. It consists of six significant modules: data acquisition, corpus extraction, data preprocessing for code-switched data, improved feature engineering module using Adaptive lemma-based vector space modeling, hybrid opinion model development, and opinion detection through multi-class classification.

The detailed components are as follows:

1) Data extraction and retrieval: At the initial stage, we retrieve online data content from heterogeneous web sources and unstructured data platforms from their respective programming APIs.

2) Opinion preprocessing: In the next phase, subtasks of NLP, including Stemming, Lemmatization, and irrelevant Noise Removal, are executed. Other subtasks may be required as per the application domain.

3) Lexicon identification: The step for detecting lexicon programmatically is crucial for efficient machine learning and deep learning model implementation. The translation and transliteration involve conversion from 1 lingual and structural format to another, enabling efficient execution of the detection phase given the heterogeneity of opinions.

4) Feature engineering: This phase comprises of feature extraction, transformation, and subsequent vectorization, is executed by employing NLP in several diversified opinion classes.

5) Evaluation: The WCM-OMF model is evaluated using standardized performance indicators depending upon subtasks for measuring the effectiveness of training and test evaluation.

The efficacy of the proposed solution is demonstrated through the comparative analysis of the proposed solution with the existing solution on benchmark dataset 1 (BD1) using standardized performance metrics. These steps are explained in detail in later sections. Further, the proposed weighted ensemble opinion mining framework (WCM-OMF) has been designed and implemented.

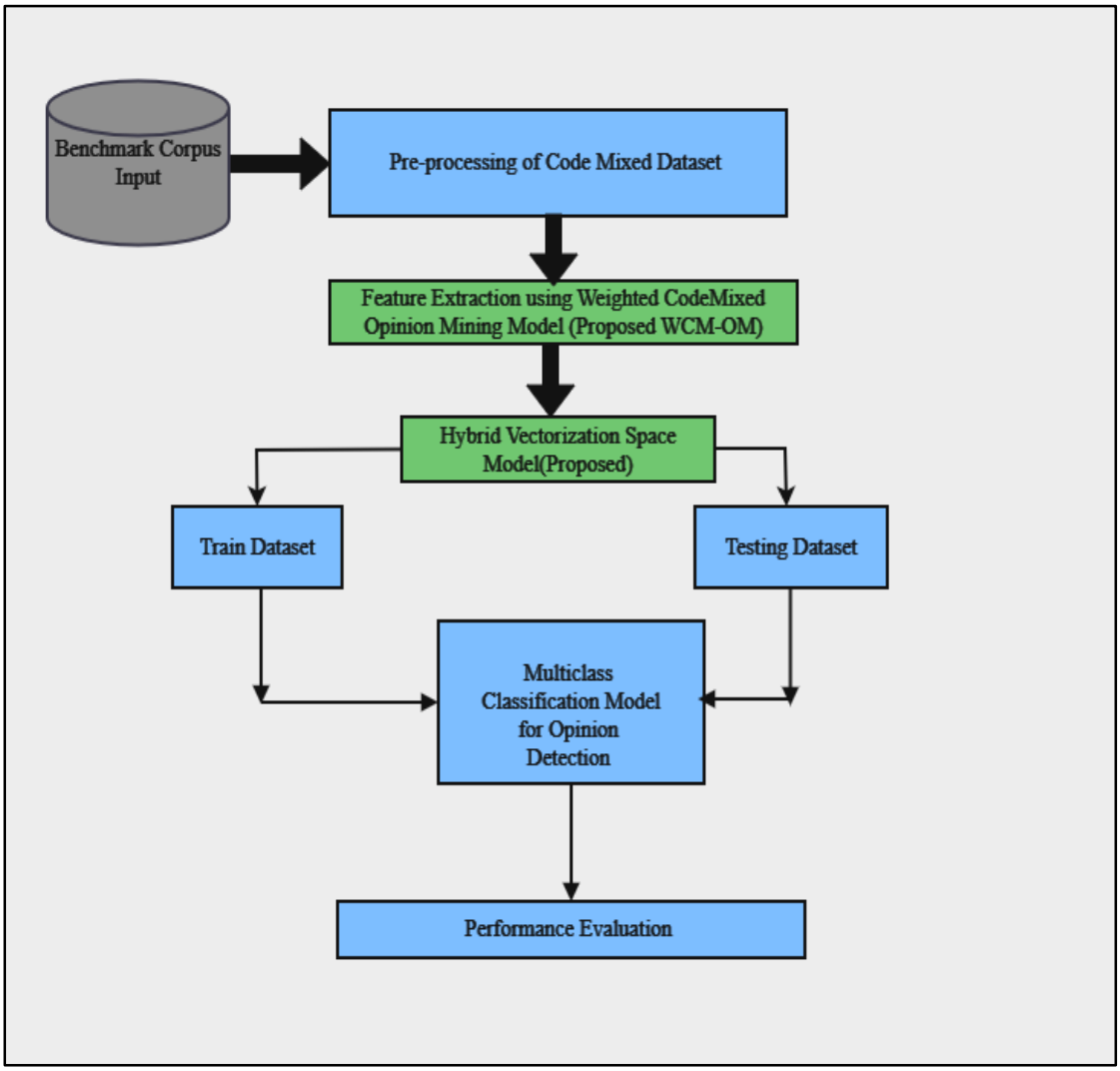


Fig. 2. Proposed opinion mining model conceptual framework.

The proposed solution assigns weights based on the following:

1) The Adaptive Bilingual Influence Weighting (ABIW)

2) The number of code-switches in the initial sentence of the opinion.

We observe that a comprehensive and holistic approach to opinion detection requires the consideration of various factors, and the current state of opinion interpretation requires further improvement consisting of the following key components:

- N: the number of tokens in a segment,
- $LangScore_i$: score indicating the relevance of the language of the i-th token,
- σ : normalization function.

Fig. 3 explains the stepwise analysis of the Proposed Solution Flowchart for the WCM-OM framework. In this architecture, the yellow highlighted cells represent the contributions of the proposed solution.

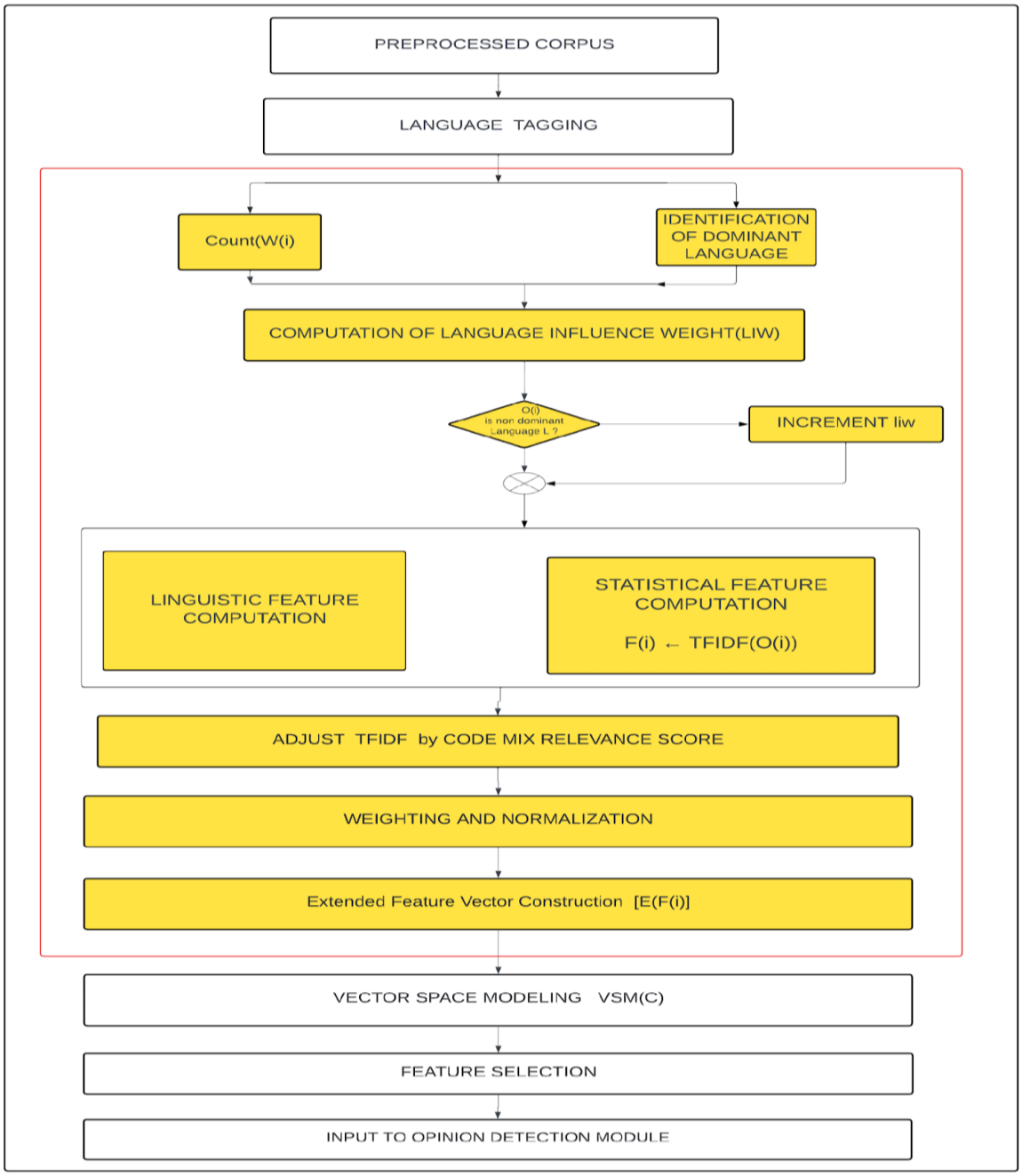


Fig. 3. Proposed WCM-OMF solution flowchart.

Firstly, the preprocessed output from the data extraction module functions as input for WCM-OMF. Thereafter, a language tagging module separates the opinion into parts of speech and tags each sub-block. Subsequently, the ABWI lemma is implemented using Language Influence Weighting (LIW) methods, a proposed metric designed to capture the influence of various lexical contexts in the code-mixed input. ABIW quantifies the extent to which the code-mixed input language contributes to the overall context. Following this, the linguistic feature is computed and combined with the statistical feature to obtain an EFV novel hybrid score. The code-mixed relevance score adjusts the TF-IDF score to get a new score. Weighting and normalization are applied to obtain the final feature vector EF.

C. Proposed Mining Method Mathematical Formulation

The equations and mathematical formulation are described in this section:

$$TF(t,\ o) = \frac{Number\ of\ times\ term\ t\ appears\ in\ a\ opinion\ o}{Total\ terms\ in\ \ opinion\ o} \quad (1)$$

$$IDF(t,\ O) = log\ \left(\frac{Overall\ opinion\ count}{Number\ of\ opinion\ with\ term\ t\ in\ it}\right) \quad (2)$$

$$\text{TF-IDF} = \text{TF (t, o)} * \text{IDF (t, O)} \quad (3)$$

Eq. (1) and Eq. (2) explain the computation of term frequency and inverse document frequencies Eq. (3) is the multiplication of term frequency, and the logarithm of opinion count to frequency ratio.

- EFV: Extended Feature Vector.
- TF-IDF: Term Frequency –Inverse Document Frequency.
- ABIW: Adaptive Bilingual Influence Weighting.

WCM_OMF is implemented to solve the code-switching problem in single opinion text (intra–sentential). For this purpose, lemma ABIW is devised as explained below.

Adaptive Bilingual Influence Weighting (ABIW)

The proposed lemma is explained in Eq. (4) that is used to dynamically adjust the influence of each language based on the opinion.

$$\text{ABIW} = \sigma\left(\frac{1}{N} * \Sigma(i = 1\ to\ N) LangScore_i\right) \quad (4)$$

Where

- W_{Lang} : weighting factor for each language in the bilingual code mixed text.
- $Model_m$: Trained model m.

Adaptive Bilingual Influence Weighting (ABIW), the language influence weighting method (LIW), is a metric designed to capture the influence of different languages in the code-mixed text. It quantifies the extent to which each language contributes to the overall content.

The weighting factor is computed for each opinion instance using a language identification function in combination with a weighting parameter. This result is added to the feature vector module, capturing the linguistic feature weights.

Further, we implement vector space modeling on the extended feature vector for the feature selection process. Eq. (5) computes the final feature vector, which is then used as modified feature vectors for multiclass opinion classification and subsequent opinion detection.

$$CMRS(O(i)) = \sum_{i=0}^{n} LF(O(i)) + STF(O(i)) \quad (5)$$

$CMRS$: Code Mixed Relevancy Score

$O(i)$: i th Opinion where i Є [1, 2, 3…n] and n is the Total count of opinions

STF –Statistical Features Set

LF- Linguistic Feature Set

Algorithm Weighted Code mixed Opinion Mining Framework (WCM_OMF)

Input Parameters:

D: Dataset with two columns,

T: Opinion text (T) and class (C),

Where T_i is the i^{th} text instance and C_i is the corresponding class label.

Algorithm : Weighted Code mixed Opinion Mining Framework(WCM_OMF)

Input: Corpus (C), Total number of samples (n).

Output : Optimized representation of Documents

Preprocessing

For each text T_i in D, preprocess the text to create a normalized text T_i'

$$T_i' \leftarrow preprocess_{text(T_i)}$$

Language Influence Weight (LIW):

Computation of LIW for each text T_i' using language id function L and weighting parameter α

$$LIW_i \leftarrow calculate_{\{liw\}(T_i', \alpha)}$$

Feature Engineering:

LIW module output conversion of T_i' into a feature vector F_i using TF-IDF

$F_i \leftarrow TFIDF(T_i')$

Combine feature vector F_i with LIW_i to create an extended feature vector EF_i

$$EF_i \leftarrow [F_i;\ LIW_i]$$

Dataset Preparation:

Class balancing module implementation ADASYN balancing on D_{train} to get D'_{train}

Training:

Train each model m in M on $D'_{\{train\}}$ using the optimized hyperparameters HP^*_m

$$Model_m \leftarrow Train(m, HP^*_m, D'_{train})$$

Development of ensemble model E from all trained models in M.

$$E \leftarrow Ensemble(M)$$

Prediction and Evaluation:

Predict classes for D_{test} using ensemble model E to get predictions P

$$P \leftarrow Predict(E, D_{test})$$

End

The proposed WCM_OMF Opinion Feature Engineering solution implements the computation of an Extended Feature Vector (EFV) instead of FV in existing systems for feature engineering opinion detection. Further, the resultant output is then inputted for embedding and classification, enabling a deeper linguistic context that assists in distinguishing opinions in different language segments.

IV Experimental Results

In this section, we present the experimental results of the proposed methodologies for code mixed data input.

Table III indicates that the proposed weighted code-mixed opinion analysis solution significantly improves opinion classification accuracy by incorporating both Hindi and English text features. The proposed solution employs a novel weighted strategy for code-switching. Quantitative and qualitative analyses are conducted for efficient data synthesis and validation. The proposed unique feature assigns higher weights for code-mixed lexicons during the opinion mining process, boosting the overall model performance. After using the k-fold cross-validation technique, the unique weighting technique's performance is further validated by performance metrics and compared to current frameworks.

Fig. 4 demonstrates the efficacy of the proposed solution for Benchmark Dataset 1 using Accuracy and Recall rate as evaluation metrics. The proposed model outperforms the existing systems SVM, Random Forest, sub-word LSTM, and Hierarchical LSTM with attention to phonetic words.

TABLE III. Comparative Analysis of Existing and Proposed Frameworks Existing Systems on Benchmark Dataset 1

Methodology	Accuracy	Recall	F1-score
Support Vector Machine Model	70.7	31.3	0.429
Random Forest Model [45]	65.1	19.3	0.292
Sub-word LSTM model [45]	69.8	36.5	0.458
Hierarchical LSTM model + attention on phonemic sub-words [47]	66.6	45.1	0.487
WCM-OMF (Proposed weighted code mixed approach)	**79**	**79.11**	**0.79102**

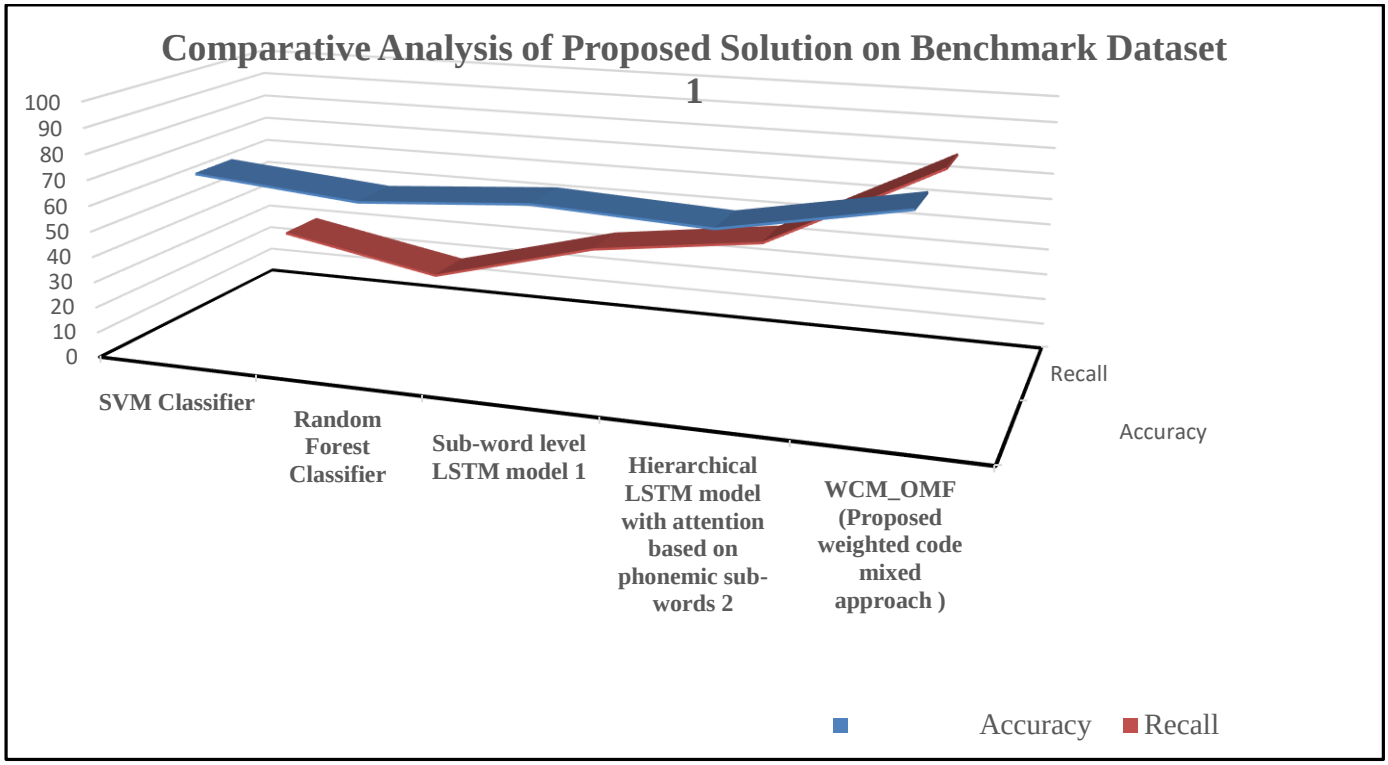


Fig. 4. Performance comparison of existing and proposed frameworks on benchmark Dataset 2.

TABLE IV. COMPARATIVE ANALYSIS OF EXISTING AND PROPOSED FRAMEWORKS ON BENCHMARK DATASET 2

Methodology	Authors	Accuracy	F1-score
BSVM (Unigram)	Wang, Manning[48]	59.15%	0.5335
NBSVM (Uni + Bigram)	Wang, Manning[48]	62.50%	0.5375
MNB (Unigram)	Wang, Manning[48]	66.75%	0.6143
MNB (Uni + Bigram)	Wang, Manning[48]	66.36%	0.6046
MNB (TFIDF)	Wang, Manning[48]	53.53%	0.4783
SVM (Unigram)	Pang and Lee[49]	57.60%	0.5232
SVM (Uni+Bigram)	Pang and Lee[49]	52.96%	0.3773
Lexicon Lookup	Sharma et al.[50]	51.15%	0.252
LSTM Model 1	Joshi et al[46]	59.80%	0.511
LSTM Model 2	Joshi et al[46]	59.70%	0.658
WCM_OMF (Proposed weighted code mixed approach)	**Proposed**	**72.55%**	**0.7208**

Results demonstrate that the proposed weighted approach to handle code-switching gives significantly better results than existing opinion mining frameworks, achieving an accuracy of 79% using comparative analysis evaluated on benchmark dataset 1. Table IV results show that the proposed weighted approach for handling code-switching performs significantly better than the existing opinion mining frameworks.

The overall accuracy achieved for the proposed system is 72.55%, and the F1 score is 0.72. The proposed method considers relevant factors during the process of grouping language-based text.

V. DISCUSSION

In this section, we discuss the results of the proposed solutions and the lemma developed. As observed in Fig. 5, the proposed solution WCM-OMF is designed and implemented to assign weights based on the ABIW lemma and the number of code switches in intra-sentential opinion. The steps for WCM (Weighted code mixing metric) are explained in detail and are primarily implemented to handle the unstructured nature of opinion input. Results show that the proposed WCM-OMF performs significantly better, achieving an F1 score of 0.72, which is a 6% improvement over the existing method on the benchmark dataset.

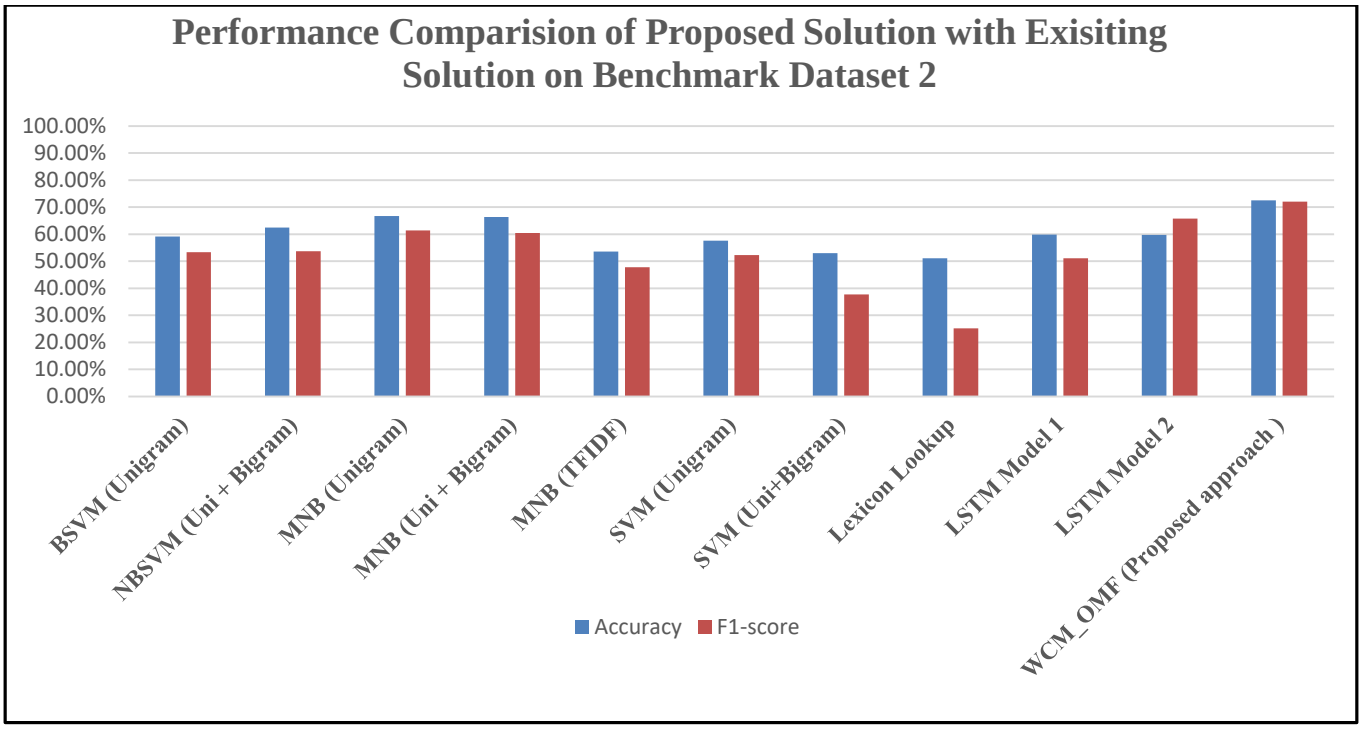


Fig. 5. Performance comparison of the proposed solution on benchmark Dataset 2.

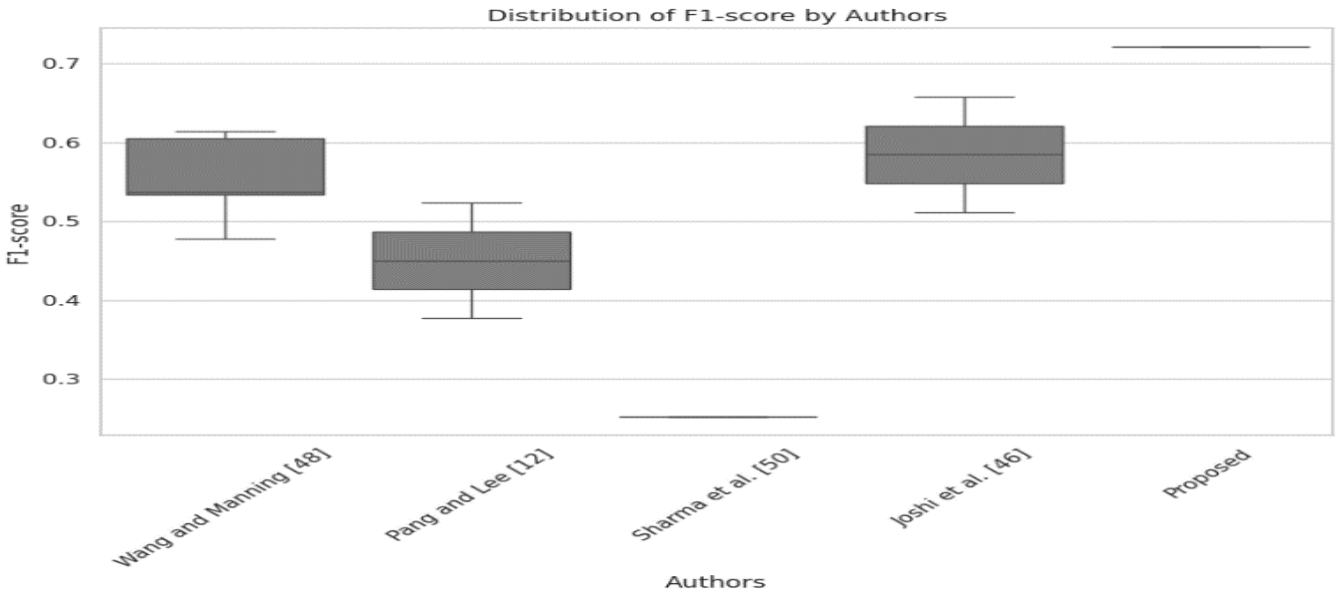


Fig. 6. Box Plot representation of F1 score distribution of proposed and existing systems.

As demonstrated in Fig. 6, the box plot represents the accuracy and F1 score distribution across several existing state-of-the-art models. As observed, the proposed model demonstrates a superior performance, as evidenced by the grouping of median data points. Opinion mining helps in the detection of patterns, improves decisions, and customizes strategy-based data extracted [51], leading to Business Intelligence [52].

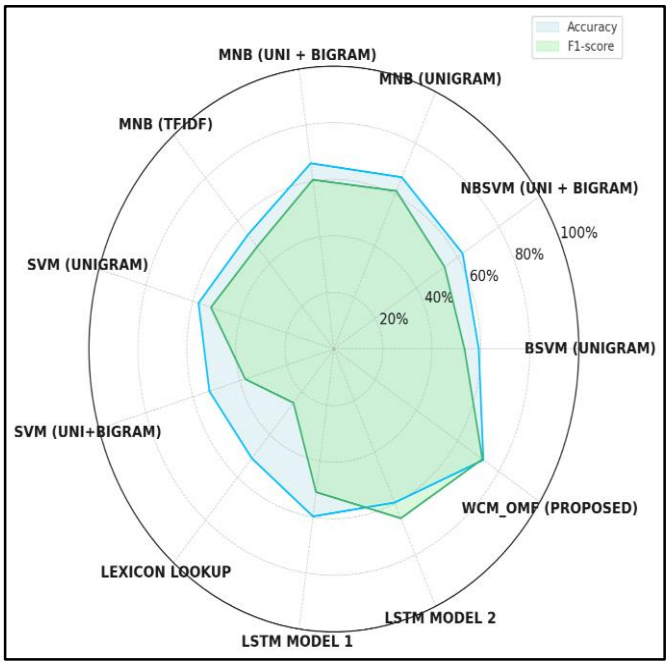


Fig. 7. Radar chart of accuracy and F1 score.

Fig. 7 visualizes the radar chart of accuracy and F1 score that demonstrates the proposed model efficiency across multiple metrics compared to the existing MNB BSVM-Unigram and BSVM- Bigram approaches.

We further validate the proposed results in Table V by implementing multiple Comparison of Means - Tukey HSD (Honestly Significant Difference) test to identify significant differences between pairs of group means with 95% confidence and p-value less than 0.05 mean difference of 9.86 for confidence interval [9.548, 10.171] and 13.21 for confidence interval [12.898, 13.521] indicating proposed method outperforms the existing approaches and differences are statistically significant.

TABLE V. TUKEYS HSD STATISTICAL SIGNIFICANCE TEST

Group 1	Group 2	Mean Diff	Lower	Upper
BSVM Unigram	MNB Unigram	7.41	7.0986	7.7214
BSVM Unigram	NBSVM Uni+Bigram	3.35	3.0386	3.6614
BSVM Unigram	Proposed	13.21	12.8986	13.5214
MNB Unigram	NBSVM Uni+Bigram	-4.06	-4.3714	-3.7486
MNB Unigram	Proposed	5.8	5.4886	6.1114
NBSVM Uni+Bigram	Proposed	9.86	9.5486	10.1714

VI. CONCLUSION

In this paper, we propose a novel WCM-OMF ensemble opinion mining framework in the context of code mixed opinions and unstructured format. We develop an end-to-end framework to handle the code-switching problem. This study also aims to bridge the gap between corpus availability and opinion mining research that would benefit researchers in analysing the proposed corpus for code mixed opinions. The results of our proposed framework are more efficient than related existing approaches in the case of code mixed opinions. The proposed method achieves an F1 score of 0.79, achieving a 6% increase in the opinion mining system. Further, we intend to extend this research to implicit opinions and word sense

disambiguation techniques concentrating on irony and other related clauses present in the opinion. This research significantly impacts creating an explainable digital environment that enhances decision-making.

Further, based on the rigorous analysis of scholarly research, we present a comprehensive framework for implementing cross-lexical opinions using cross-platform resources. Further, we suggest future developmental efforts, namely:

- Low resource corpus availability.
- Multimodal data integration capabilities.
- Open source platforms for handling big data analytics.
- Optimizing resource consumption for transformer architectures and deep learning methodologies.

The proposed solution would enable the utilization of potential applications for enterprises having an online presence for effective decision-making and business intelligence.

ACKNOWLEDGMENT

We are grateful for the insightful comments given by anonymous reviewers during the drafting of the paper.

REFERENCES

[1] Pang B, Lee L et al (2008) Opinion mining and sentiment analysis. Found Trends Inf Retr 2(1–2):1–135.

[2] Tong RM (2001) An operational system for detecting and tracking opinions in on-line discussion. In: working notes of the ACM SIGIR 2001 workshop on operational text classification, vol. 1

[3] P. Liang and B. Dai, "Opinion mining on social media data", 14th International Conference on Mobile Data Management, pp. 91-96, 03-06 June 2013, https://doi.org/10.1109/MDM.2013.73.,IEEE

[4] C.C. Aggarwal and C.X. Zhai (eds.), Mining Text Data, doi: 10.1007/978-1-4614-3223-4_13, Springer Science Business Media, LLC, 2012, Springer

[5] C. M. Scotton, "The possibility of code-switching: motivation for maintaining multilingualism," Anthropological linguistics, pp. 432–444, 1982

[6] Zhang, Chenxi, and Zeshui Xu. "Gaining Insights for Service Improvement through Unstructured Text from Online Reviews." Journal of Retailing and Consumer Services 80, (2024): 103898. https://doi.org/10.1016/j.jretconser.2024.103898

[7] H. Liu, I. Chatterjee, M. Zhou, X. S. Lu and A. Abusorrah, "Aspect-based sentiment analysis: A survey of deep learning methods", *IEEE Trans. Comput. Soc. Syst.*, vol. 7, no. 6, pp. 1358-1375, Dec. 2020

[8] Medhat Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment Analysis Algorithms and Applications: A Survey." Ain Shams Engineering Journal 5, no. 4 (2014): 1093-1113. https://doi.org/10.1016/j.asej.2014.04.011

[9] Andrés Montoyo, Patricio Martínez-Barco, Alexandra Balahur, Subjectivity and sentiment analysis: an overview of the current state of the area and envisaged developments,Decis Support Syst, 53 (2012), pp. 675-679

[10] Date, Saroj S., Mahesh B. Shelke, Kiran V. Sonkamble, and Sachin N. Deshmukh. "A Systematic Survey on Text-based Dimensional Sentiment Analysis: Advancements, Challenges, and Future Directions." Computational Intelligence Methods for Sentiment Analysis in Natural Language Processing Applications, (2023): 39-57. https://doi.org/10.1016/B978-0-443-22009-8.00014-8.

[11] Yoon, Joanne. 'Classifying Respondent Comments from the 2021 Canadian Census of Population Using Machine Learning Methods [1]'. 1 Jan. 2023: 785 – 791.

[12] Laifa, M. and Mohdeb, D. (2023), "Sentiment analysis of the Algerian social movement inception", Data Technologies and Applications, Vol. 57 No. 5, pp. 734-755. https://doi.org/10.1108/DTA-10-2022-0406

[13] Sietsma, A.J., Groenendijk, R.W. & Biesbroek, R. Progress on climate action: a multilingual machine learning analysis of the global stocktake. Climatic Change 176, 173 (2023). https://doi.org/10.1007/s10584-023-03649-3

[14] Kohei Watanabe (2021) Latent Semantic Scaling: A Semisupervised Text Analysis Technique for New Domains and Languages, Communication Methods and Measures, 15:2, 81-102

[15] Regazzi, L., Cadeddu, C., Gris, A. V., Lomazzi, M. (2023). Sentiment towards vaccinations in Chinese healthcare workers: preliminary results from an international survey. Population Medicine, 5(Supplement), A2051. https://doi.org/10.18332/popmed/165050

[16] Schrader, Samuel R., and Eren Gultepe. "Analyzing Indo-European Language Similarities Using Document Vectors." Informatics 10, no. 4 (2023): 76. https://doi.org/10.3390/informatics10040076.

[17] X. Zhang, J. Zhao, Y. LeCun, Character-level convolutional networks for text classification, Adv. Neural Inf. Process. Syst. 28 (2015).

[18] J. Du, L. Gui, R. Xu, Y. He, A Convolutional Attention Model for Text Classification," Natural Language Processing and Chinese Computing, 2018, pp. 183–195, https://doi.org/10.1007/978-3-319-73618-1_16

[19] K. Kumari, S.S. Jha, Z.K. Dayanand, and P. Sharma, "ML&AI IIITRanchi@DravidianLangTech: Fine-Tuning of Indic-BERT for Exploring Language-Specific Features for Sentiment Classification in Code-Mixed Dravidian Language," DravidianLangTech 2023 - 3rd Workshop on Speech and Language Technologies for Dravidian Languages, associated with 14th International Conference on Recent Advances in Natural Language Processing, RANLP 2023 - Proceedings, pp. 192-197, 2023, doi: 10.26615/978-954-452-085-4_027.

[20] K. Maity, T. Sen, S. Saha, and P. Bhattacharyya, "MTBullyGNN: A Graph Neural Network-Based Multitask Framework for Cyberbullying Detection," IEEE Transactions on Computational Social Systems, vol. 11, no. 1, pp. 849-858, 2024, doi: 10.1109/TCSS.2022.3230974.

[21] R. Pandey and J. P. Singh, "BERT-LSTM model for sarcasm detection in code-mixed social media post," Journal of Intelligent Information Systems, vol. 60, no. 1, pp. 235-254, 2023, doi: 10.1007/s10844-022-00755-z.

[22] R.R. Frias, R.P. Medina, and A.S. Sison, "Attention-based Bilateral LSTM-CNN for the Sentiment Analysis of Code-mixed Filipino-English Social Media Texts," 2023 International Conference on Digital Applications, Transformation and Economy, ICDATE 2023, doi: 10.1109/ICDATE58146.2023.10248926.

[23] Nascimento, Francimaria R., George D. Cavalcanti, and Márjory Da Costa-Abreu. "Unintended Bias Evaluation: An Analysis of Hate Speech Detection and Gender Bias Mitigation on Social Media Using Ensemble Learning." Expert Systems with Applications 201, (2022): 117032. https://doi.org/10.1016/j.eswa.2022.117032.

[24] Manikyamba, I. Lakshmi, and A. Krishna Mohan. "Using a Novel Hybrid Krill Herd and Bat based Recurrent Replica to Estimate the Sentiment Values of Twitter based Political Data." (2023).

[25] Kar, Purbani, and Swapan Debbarma. "Sentimental Analysis & Hate Speech Detection on English and German Text Collected from Social Media Platforms Using Optimal Feature Extraction and Hybrid Diagonal Gated Recurrent Neural Network." Engineering Applications of Artificial Intelligence 126, (2023): 107143. . https://doi.org/10.1016/j.engappai.2023.107143.

[26] M. Alhawarat and A. O. Aseeri, "A Superior Arabic Text Categorization Deep Model (SATCDM)," in IEEE Access, vol. 8, pp. 24653-24661, 2020, doi: 10.1109/ACCESS.2020.2970504.

[27] Md Ashik, A. U.-Z, S. Shovon, S. Haque, Data set for sentiment analysis on Bengali news comments and its baseline evaluation, 2019 International Conference on Bangla Speech and Language Processing (ICBSLP) (2019)

[28] Md. Arid Hasan, "Ensemble Language Models for Multilingual Sentiment Analysis," University of New Brunswick, 2024.

[29] I. Javed and H. Afzal, "Opinion Analysis of Bi-Lingual Event Data from Social Networks," in Proceedings of the Conference on Language Resources and Evaluation, pp. 3172-3179, 2014.

[30] Hashmi, E., Yayilgan, S.Y. & Shaikh, S. Augmenting sentiment prediction capabilities for code-mixed tweets with multilingual transformers. Soc. Netw. Anal. Min. 14, 86 (2024). https://doi.org/10.1007/s13278-024-01245-6

[31] Altaf A, Anwar MW, Jamal MH, et al (2023) Exploiting linguistic features for efective sentence-level sentiment analysis in urdu language. Multimedia Tools and Applications pp 1–27

[32] Ghasemi, R., Ashrafi Asli, S. A., & Momtazi, S. (2022). Deep Persian sentiment analysis: Cross-lingual training for low-resource languages. Journal of Information Science, 48(4), 449-462. https://doi.org/10.1177/0165551520962781

[33] S. Thara and P. Poornachandran, ‘‘Social media text analytics of Malayalam–English code-mixed using deep learning,’’ J. Big Data, vol. 9, no. 1, pp. 1–25, Dec. 2022.

[34] Soufian Jebbara and Philipp Cimiano. 2019. Zero-shot cross-lingual opinion target extraction. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics, 2486–2495. DOI:https: //doi.org/10.18653/v1/N19-1257

[35] Ignacio González Godino and Luis Fernando DHaro.Gth-upm at tass: Sentiment analysis of tweets for spanish variants. Proceedings of TASS, 2019

[36] Michael Neumann and N. goc Thang Vu. 2018. Cross-lingual and multilingual speech emotion recognition on English and French. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 5769–5773. DOI:https://doi.org/10.1109/ICASSP.2018.8462162

[37] Sujata Rani and Parteek Kumar. Deep learning based sentiment analysis using convolution neural network. Arabian Journal for Science and Engineering, 44(4):3305–3314, 2019.

[38] Santosh T. and K. Aravind (2019). Hate speech detection in hindi-english code-mixed social media text. In Proceedings of the ACM India Joint International Conference on Data Science and Management of Data, pp. 310–313

[39] Saumya, Sunil, Abhinav Kumar, and Jyoti P. Singh. "Filtering Offensive Language from Multilingual Social Media Contents: A Deep Learning Approach." Engineering Applications of Artificial Intelligence 133, (2024): 108159. https://doi.org/10.1016/j.engappai.2024.108159.

[40] Tasnia, R., Ayman, N., Sultana, A. et al. Exploiting stacked embeddings with LSTM for multilingual humor and irony detection. Soc. Netw. Anal. Min. 13, 43 (2023). https://doi.org/10.1007/s13278-023-01049-0

[41] Mandl , T., Modha, S., Kumar M. A., & Chakravarthi, B. R. (2020). Overview of the HASOC track at fire 2020: Hate speech and offensive language identification in Tamil, Malayalam, Hindi, English and German. In Proceedings of the 12th annual meeting of the forum for information retrieval evaluation (pp. 29–32).

[42] Khan L, Amjad A, Afaq KM et al (2022) Deep sentiment analysis using cnn-lstm architecture of English and roman urdu text shared in social media. Appl Sci 12(5):2694

[43] Yang, Zijian Győző, Laki, László János (2023) Solving Hungarian natural language processing tasks with multilingual generative models Annales Mathematicae et Informaticae. 57. pp. 92-106. ISSN 1787-6117 (Online)

[44] Lany L. Maceda, Arlene A. Satuito and Mideth B. Abisado, “Sentiment Analysis of Code-mixed Social Media Data on Philippine UAQTE using Fine-tuned mBERT Model” International Journal of Advanced Computer Science and Applications(IJACSA), 14(7), 2023. http://dx.doi.org/10.14569/IJACSA.2023.0140777

[45] Bohra, A., Vijay, D., Singh, V., Akhtar, S.S., & Shrivastava, M.,” A dataset of Hindi–English code-mixed social media text for hate speech detection” In second workshop on computational modeling of people’s opinions, personality, and emotions in social media, Association for Computational Linguistics (ACL)

[46] Joshi, A., Prabhu, A., Shrivastava, M., and Varma, V. Towards sub-word level compositions for sentiment analysis of Hindi-English code mixed text, COLING, pages 2482–2491, ACL Anthology,2016

[47] Santosh, T.Y.S.S., Aravind, K.V.S.: Hate speech detection in Hindi-English code-mixed social media text. ACM Int. Conf. Proc. Ser. (2019)

[48] Sida Wang and Christopher D. Manning. 2012. Baselines and bigrams: Simple, good sentiment and topic classification. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2, ACL ’12, pages 90–94, Stroudsburg, PA, USA. Association for Computational Linguistics, ACM

[49] Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. Found. Trends Inf. Retr., 2(1-2):1–135

[50] Shashank Sharma, PYKL Srinivas, and Rakesh Chandra Balabantaray, Text normalization of code mix and sentiment analysis. In 2015 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2015, August 10-13, 2015, pages 1468–1473.

[51] S. Ruchi and S. Pravin, “Improved opinion mining for unstructured data using machine learning enabling business intelligence,” Journal of Advances in Information Technology, vol. 14, no. 1, 2023.

[52] R. Sharma and P. Srinath, "Business Intelligence using Machine Learning and Data Mining techniques - An analysis," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2018, pp. 1473-1478, doi: 10.1109/ICECA.2018.8474847